PREDICTING PILE CONSTRUCTION PRODUCTIVITY LOSS USING MACRO IMPACT FACTORS IN INDONESIA

Thesis submitted to National Central University in partial fulfillment of the requirements for the degree of master

DUAL DEGREE PROGRAM
BRAWIJAYA UNIVERSITY AND NATIONAL CENTRAL UNIVERSITY, TAIWAN, ROC
GRADUATE INSTITUTE OF CONSTRUCTION ENGINEERING AND MANAGEMENT

By:
DIAN PRAMITA SARIE
156060112111001

MASTER PROGRAM OF FACULTY OF ENGINEERING
BRAWIJAYA UNIVERSITY
MALANG
2017

PREDICTING PILE CONSTRUCTION PRODUCTIVITY LOSS USING MACRO IMPACT FACTORS IN INDONESIA

THESIS

MASTER PROGRAM OF CIVIL ENGINEERING

FIELD : CONSTRUCTION ENGINEERING AND MANAGEMENT

By :

DIAN PRAMITA SARIE

156060112111001

MINISTRY OF RESEARCH, TECHNOLOGY, AND HIGHER EDUCATION

BRAWIJAYA UNIVERSITY FACULTY OF ENGINEERING

MASTER AND DOCTORAL PROGRAM

MALANG, INDONESIA

2017

ABSTRACT

Pile construction productivity loss in Indonesia had been occurred for years. Before improving pile construction productivity, impact factors and how much potential loss are urgent to identified. The research objectives are to identify the macro factors that influence pile construction, to develop a SVR model that precisely predicts productivity loss, and to provide potential loss quantities using the most similar historical case(s). Literature review identifies 5 macro factors (labor, management, environment, material, and equipment) and 8 inputs (soil condition, pile type, pile material, project size, project location, pile depth, pile quantity, and equipment quantity) for Support Vector Regression (SVR) model, and then leads the study to collect 110 pile construction projects among 5 major areas in Java island of Indonesia. The SVR evaluated using 10-way cross validation yields an accuracy rate at 87.2%. The most likely productivity loss obtained based on the most similar historical cases is approximately 18.55% of total productivity. The findings would push the practitioners to pay attention to the loss in order to improve the overall productivity.

Keywords : *pile construction, macro factor, SVR, productivity loss.*

ACKNOWLEDGEMENT

I am grateful to Allah SWT for His blessing in completing this thesis. I would like to express my sincere appreciation to my advisor, Professor Jieh-Haur Chen, for his knowledge, guidance, patience, and constant motivation to finish my thesis. I am also grateful to all the professors and staffs in Graduate Institute of Construction Engineering and Management in National Central University Taiwan, for their advices, knowledge, and support during one year I studied here.

My sincere gratitude to my advisor in Indonesia, Dr. M. Ruslin Anwaar, M.Si., for his advice and support, although we were far away. I would also like to acknowledge all staffs in Universitas Brawijaya, Indonesia, and all of committee members Prof. Ying-Hua Huang 黃盈樺 and Prof. Shu – Chien Xi 徐書謙 for excellent comments and suggestions.

I would also like to acknowledge all of Indonesian Contracting Companies participated in this study, especially for Rudi Setyawan (as my fiancée) for his moral support and encouragement, as well as everyone who provided me with relevant information for this research.

I am very grateful to all of my lab mates, both in my 1st semester and 2nd semester for their friendship. My sincere thanks to all members of Purple Bull Team, especially for Mia, Eric, Vidya, Man-Nhi, and Sam for their help, support, for great time and experiences with them.

Sincere thanks to my Indonesian friends, PPI NCU, Alumni ICD 2017 team members, and DD Sipil thank you for being my second family. Especially for Manis Manja Group (Alfinna, Raras, Ruri, Aghnia, Annisa), Punakawan (Ganis, Itong, Samid, Nyox, Angga), Dewi, Niken, and Ario, thank you for your continuous support although we were in different countries.

Finally, my deepest acknowledgment is dedicated to my beloved parents, Mr. Bambang and Mrs. Dyah, my sisters Arda, Intan and Nata The Little Angel for supporting me, their love, prayer and wishes the best for me. Thank you for being my motivation when I lived far away from home.

~Dian Pramita Sarie~

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I : INTRODUCTION

1.1 Research Background

The construction of pile foundation is complicated by enormous number of problems relating to several factors such as subsurface obstacles, contractor's lack of experience, and site planning difficulties. Due to unknown site condition, productivity loss is inevitable for pile construction projects. Site pre-investigation usually consists of statistical samples around foundation area that did not cover the entire area. Soil types differ from site to site due to soil cohesion or stiffness, natural obstacles, and subsurface infrastructure construction obstacles. Lack of experience in adjusting the pile axis, length, and size also present further complication. Besides, piling machine mechanical and drilling problems also need to be considered. Problems due to site restrictions and disposal of excavated spoil have a great effect on productivity. The rate of rebar installation and concrete pouring affected by rebar crews' experience and pouring method. All these problems, no doubt, greatly affect the production of concrete bored piles on site. Therefore, the objectives of current study are to identify and measure the macro impact factor for pile construction productivity in Indonesia, also to analyze pile construction productivity factors using the regression analysis technique based upon the mentioned factors. In addition, it provides researchers with the methodology of designing regression models for pile construction process, its limitations, and future suggestions.

Indonesian construction sector is developing rapidly with both private and public projects on the rise. Housing backlog and property sector growth in several big cities within the country caused construction sector in Indonesia to grow 7-8 percent (Bank, 2009). Public works investment is key in the government plan to deliver roads, water resources and human settlement infrastructure for long-term development.

According to the publication *Asia Construction Outlook,* Indonesia is the one of the top-rated country in terms of potential construction spending growth in the medium-term. Indonesia's construction market totaled US$ 49.2 billion in 2014 (AECOM). Construction spending is expected to grow about 5.2% per year in the 2014-2019 period, which is above the regional average of 4.4 percent (AECOM). In fact, the publication identified Indonesia as the second most profitable construction market in Asia (AECOM). The country also has increasingly open and attractive market for investment.

Methods for measuring construction productivity remains unavailable, so it is difficult to perform project control (e.g management technical aspects). Most of the construction industries measure their pile productivity by calculating the delays from project deadline. The above – mentioned major problems create difficulties for the estimator to evaluate pile construction productivity and find the factors that affect on-site productivity the most.

1.2 Research Objectives

Evaluating productivity is a rather complicated problem, for almost all the factors involved in a project eventually affect project's productivity. Many existing researches dealt with influential factors of productivity in a more qualitative manner; important factors were simply identified but not quantified, nor were relationships with other factors established. The present study qualitatively and quantitatively identifies critical factors leading to the loss of productivity, and the cause-effect relationship between these critical factors, and allowing the project organizational-level factors to be determined. Findings from this research can provide managers with some guidelines and strategy to help them efficiently plan and execute their projects in such a way as to enhance productivity. The followings are the objectives of the study:

a. Identify macro factors that impact the pile productivity. Afterward, the extent of the identified factors' effect towards the project-level productivity will be measured using regression model framework for concrete pile construction
b. Develop Support Vector Regression model in order to predict productivity loss in Indonesian pile construction productivity.

1.3 Research Scope and Limitations

The scopes of this study are set to: (1) pile productivity for construction building project; (2) Project data derived from Indonesia; (3) Computational prediction using regression method; and (4) Private sector project cases.

1.4 Methodology

The research was conducted through a combination of literature review, personal interview with pile construction companies, as well as expert judgement regarding factors that affect pile construction productivity. Collected data includes qualitative and quantitative data sets and analyzed data. The first few months' literature review provided a basic recognition of influential factors of construction productivity, resulting on a basic framework for the research.

After conducting several interviews with the professionals, the loss of working hour onsite were determined, and a carefully designed questionnaire was sent out to eight ongoing projects to get the information on work and project performance.

After acquiring enough data and information, a model of significant influential wastes for construction productivity was developed as a regression model. Then the causes for these significant wastes were traced to the project level. The last step on this research is to conduct a case study to serve the purpose of examining macro impact factors through pertinent examples that provide some evidence of similar results thus allowing analytical generalizations to be made.

1.5 Research Flowchart

The flowchart showed in Figure 1 illustrates steps used in performing this study. First, background and problem statements demonstrate the importance of this research based on previous studies. Second, collection of some factors as attributes from literature review to analyze the significant ones. Third, the data were collected from construction projects in Indonesia. After data collection, factor analysis was performed to determine the rank of significant factors.

The data collected were analyzed by standardizing the content of each attribute before prediction. We used the R software for prediction analyses. After the value of accuracy is accepted, the proposed model could be built.

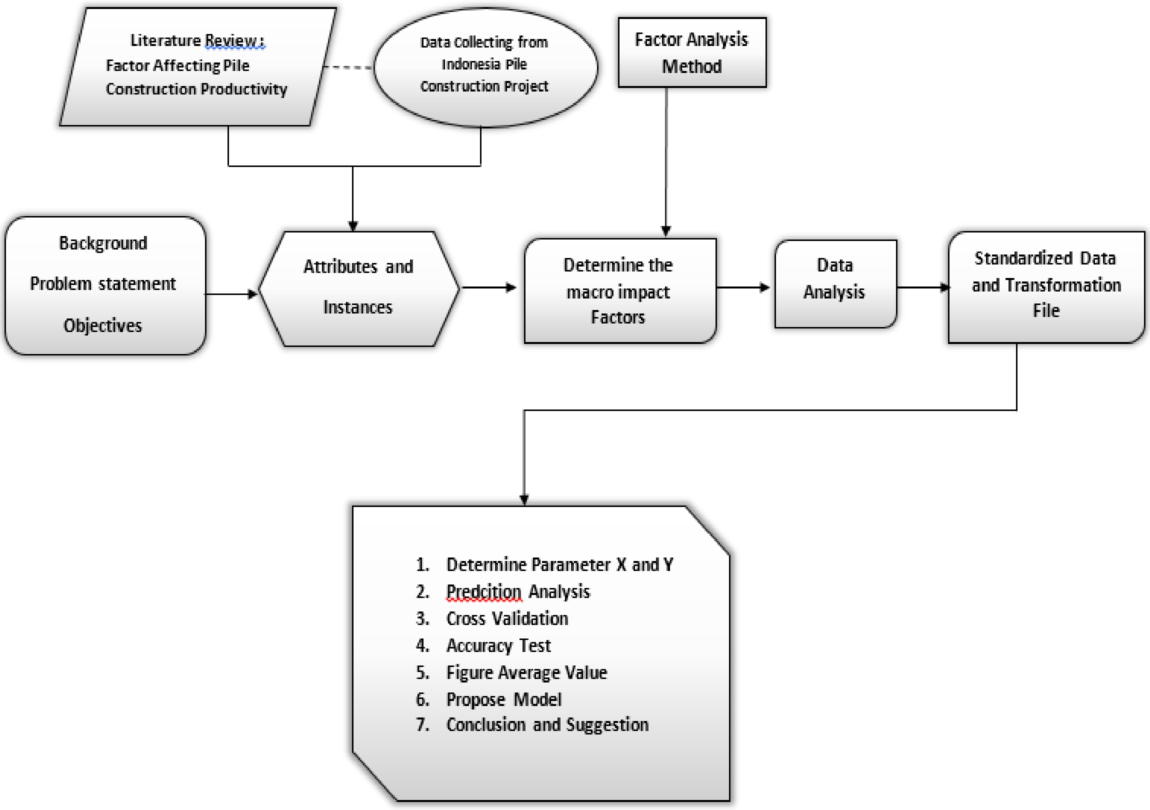


Figure 1.1 Research Flow Chart

1.6 Thesis Organization

The study consists of five chapters including an introduction, literature review, data collection and analysis, model development, and conclusion.

Chapter I: Introduction, describes about the research background and problem statement, objectives of the study, research scope and limitations, research flowchart, and thesis organization.

Chapter II: Literature Review, provides the theory of pile construction productivity prediction, prediction accuracy, macro impact factor affecting the pile construction productivity, and Support Vector Machines (SVM).

Chapter III: Data Collection and Analysis, comprises the collected data and statistical analysis.

Chapter IV: Model Development, examines the process and development of SVM regression.

Chapter V: Conclusions, reported the results and findings in each region of study target. It contains research summary, conclusion, and also discusses future research prospects in pile construction productivity in the project.

CHAPTER II : LITERATURE REVIEW

2.1 Productivity Concept

In simple terms, productivity can be expressed as the relationship between the output generated from a system and the input used to create output. Inputs generally refer to labor, capital, energy and materials which are brought into a system. These resources are transformed into outputs, i.e. goods and services. There are basic forms of productivity, namely:

• Total productivity (also known as total factor productivity)

• Partial productivity (also known as partial factor productivity).

Total productivity is the ratio of total output to all input factors. Partial productivity is the ratio of total output to one class of input, e.g. labor productivity is a partial productivity measure. The equation for total and partial productivity can be seen on equation (1) and (2):

$$\textit{Total productivity} = \frac{\textit{Total Output}}{\textit{Total Input}} \tag{1}$$

$$\textit{Partial Productivity} = \frac{\textit{Total Output}}{\textit{Partial Input}} \tag{2}$$

Based on the equation, when all outputs and inputs are included in the productivity measure it is called total productivity. Total productivity is a measure of total efficiency of a production process and as such is often presented as the objective to be maximized in a production process such as a in construction project (Sumanth). The measurement of productivity shall be developed so that it will indicate increases or decreases in the productivity of the project.

The major difficulties in applying the above simple productivity equations center around:

a. Measuring output, especially with regard to changes with time in the sizes and types of individual products
b. Measuring inputs and accounting for the great diversity in types of materials, facilities and equipment needed as well as the multiplicity of labor skills to be encompassed
c. Determining which particular input-output comparisons are most relevant in evaluating the performance of various operations of concern to management
d. Interpreting productivity figures in order to differentiate between the influences of internal and external factors.

In spite of the difficulties encountered in measuring productivity both industrialists and researchers alike believe in the advantages that it can bring four reasons as to why it is necessary to measure productivity:

a. *for strategic purposes,* in order to compare the performance of the firm with that of its competitors or related firms, both in terms of aggregate results and in terms of major components of performance;
b. *for tactical purposes,* to enable management to control the performance of the firm by identifying the comparative performance of individual sectors of the firm, either by function or by product;
c. *for planning purposes,* to compare the relative benefits accruing from the use of different inputs, or varying proportions of the same inputs, currently and over longer periods, as the basis for considering alternative adjustments over future periods
d. *for other management purposes,* such as collective bargaining with trade unions.

2.2 Pile Construction Productivity Measurement and Performance

Pile Construction productivity measurement is an analysis into the ratio of total output versus the total input of the construction process (Mawdesley M J, 2000). Thomas and Mathews in Park (H.-S. Park, Thomas, S.R. & Tucker, R.L., 2005) stated that no standardized productivity definition has been established in the construction industry. From the 1960s, developed countries have grappled with the problem of productivity in the construction industry with a view to understanding the basic problem of how to measure it (Jan Bröchner, 2013). However the definition of productivity itself has had its fair share in the confusion prevalent in the computation of productivity data. For example many companies measure profitability and report it as productivity. The difference between these two is that whereas profitability is the monetary process, productivity is a physical process (Pekuri, Haapasalo, & Herrala, 2011). However, productivity is eventually measured in monetary units of output per level of input (Stainer, 1997). Profitability has been known to change for reasons unrelated to productivity hence the base measures of productivity remain crucial. Productivity has also been mistaken with performance.

The measurement problem is exacerbated by the fact that the construction industry is composed of four sectors that differ significantly in; the outputs produced, firm size, and use of technology. The four sectors, which taken together define the construction industry, are residential, commercial/institutional, industrial, and infrastructure (Huang, 2009). This means that for each of these unique sectors, which in most cases construction companies are engaged in simultaneously, the contractor has to develop systems of measuring the different forms of output and the varied conditions under which all these projects are carried out.

Huang (Huang, 2009) proposed what became the major delimitation criteria for this study when they identified that the construction industry has three distinct levels as far as productivity is concerned; task level, project level and industry level. *Task* refers to a specific activity like pouring concrete or structural steel erection, *projects* refers to the collection of activities which result in the renovation or the construction of a facility. *Industry* refers to the total portfolio of projects within an economy.

Jonssons (Jonsson, 1996) highlighted that the construction company was an entity that had various projects where the actual productivity was taking place. This then places the projects in a key position with regards to productivity measurement as the ultimate products of construction. With this in mind this study will focus on project level productivity because according to Park (H.-S. Park, 2006), Construction productivity rates differ between projects because of the varying environments, characteristics, and project management efforts for each project. Therefore, when analysing construction productivity, one should consider the drivers that cause construction productivity differences between projects.

Chan and Kaka (Paul W. Chan, 2007) took the case for project based measurement of productivity further by highlighting that although strategic levels of management were crucial in improving construction productivity, there was a need to relate it to the projects themselves. They quote Groak (Groak, 1994) who reinforces this by saying that industry had gone amiss by failing to recognise the project site as the “defining locus of production organisation”. This meant that the industry needed to reorient its production improving efforts to the projects, and to do this measurement becomes critical.

2.3 Pile Construction Productivity Measurement Methods

There is information needed to calculate a meaningful project level productivity metric. For instance, information yielding the task weight (share that it represents to the overall project) is required, as is an understanding of the task flows. Because some tasks are completed in parallel, while others in series, the composition of the task flows affects overall project productivity. Therefore, each component of the project productivity metric contains: (1) the task weight; (2) the raw task productivity baseline value in the denominator; (3) the raw task productivity value for that project in the numerator; and (4) a measure of the task mix (in parallel versus in series task flows). The project productivity index value is a function of the individual components (Industry, 2014)

Huang et al (Huang, 2009) also proposed that an alternative project level productivity index can be produced as follows. We can create an index which is the quotient of two ratios, in each ratio the numerator is the value of construction put in place and the denominator is the number of field work hours. As noted earlier, a reference data set can be used to fix a baseline value for the ratio of value put in place to field work hours. The baseline value for the ratio is then used as the denominator in the index calculation. How an individual project compares to the baseline is determined by inserting its ratio of value put in place to field work hours in the numerator of the index.

2.4 Multiple Regression Model Coeficcient Pile Construction

The effect magnitude obtained by using regression analysis where the dependent variable is the big productivity difference in basic and actual daily productivity, while the independent variables are value findings, recording the factors happens that environment (X_1), equipment (X_2), labor (X_3), materials (X_4), and management (X_5) factors.

In dependent variables calculation procedure, if the result of base value reduced by actual daily productivity shows negative number, thus, 0 should be inserted because it shows good result where the actual productivity is higher than ideal productivity.

$$\boldsymbol{Base - Act\ Productivity} = 0{,}11\ X_1 +\ 0{,}89\ X_2 +\ 0{,}51\ X_3\ + 0{,}09\ X_4\ + 0{,}77\ X_5 \qquad \textbf{(3)}$$

Source : (Lesmana., 2014)

Loss of productivity is the value of model coefficient times binary factors that occurs on each working day during the research, while *workhours lost* is lost working hours due to different factors that occurred on that working day. The value of *workhours lost* is the value of *loss of productivity* times with working hour (Lesmana., 2014).

In the case of Indonesian pile construction projects, especially in Java Island, factor that has the biggest impact was equipment because there are damage on the equipment, thus causing idle time that greatly reduced productivity.(Lesmana., 2014) Lesmana (Lesmana., 2014) states that the second factor that greatly reduced productivity was labor, because the labors often take breaks outside the provided break time, thus causing *idle time*.

2.4.1 Average Productivity in Pile Construction Productivity

The minimum, maximum, range and the average productivity rates for all the trades under observation and interview indicated large variation of productivity rates over sites and generally supported the fact that baseline productivity rate attached to an activity cannot be fixed, as there are several factors interacting with each other, affecting the overall productivity. The productivity figures also differed significantly with the existing database of productivity rates of the case study company, concurring with the results of Olomolaiye (Olomolaiye P, 1998).

The reasons for this difference were attributed to technical problems associated with construction trades, based on the location of the site, soil strata, contract specifications and client involvement, besides the factor variables considered in the study.

This score is expected productivity when the factors occurred, which shows the result of deducting model coefficient from average productivity. Besides, the scores also represent work hour efficiency when the factors occurred. The amount is expected unit rate divided by average productivity, on other words, the ration between expected productivity and baseline productivity. (Lesmana., 2014)

Table 2.1 Sample of Productivity Average Calculations Pile Bored Piling Works

No	Date	Pile Depth (A)	Number of pile finish/day (B)	Total Dept (A x B)	Work Hour / day (C)	Equipment sheet (D)	Total Work Hour of Equipment (C x D)	Daily Productivity (A x B / C x D)	Average Productivity
		m	qty	m	hour	qty	hour	m / hour	m / hour
4	9/18/2014	52	7	364	18	4	72	5.06	5.51
5	9/19/2014	52	5	260	18	3	45	5.78	
6	9/20/2014	52	2	104	18	2	20	5.20	
7	9/21/2014	42	3	126	18	2	20	6.30	
10	9/24/2014	52	3	156	18	2	30	5.20	

Source: Project Data, 2016

2.4.2 Impact Factors of Pile Construction Productivity Measurement on Performance

For a contractor, field productivity represents the single biggest risk and differentiating factor when it comes to project execution. This implies that assessing the productivity on a project can greatly increase contractors' awareness of this risk and help them mitigate it. Winch and Carr (Winch, 2001) gave a shocking industry study when they realized that of the companies that they studied, the very fact that they were doing a productivity study led the workforce to increase their productivity rates. In other words, the competitiveness and drive to please increased when they perceived that their work was being held to some kind of standard.

The first to note that worker morale was boosted by noting their daily accomplishments when he was making his case for task recording time cards. Highlighted that worker morale was one key resultant from productivity improvement, contactors who measured their productivity were better placed to improve it and hence they experienced a greater morale on their projects. Conversely he discovered that poor worker morale led to poorer productivity.

Santosh and Apte (Varma Santosh, 2014) in studying productivity measurement also noted that the labor force was motivated by getting feedback on their performance. They found a greater performance was achieved because the workers were aiming at set targets. They went on to list more benefits that they found were accrued to contractors from performing productivity measurement, namely;

a. Decreased total cost of production
b. Decreased total duration of production

The coefficient value of each factor is a value which represents the influence of each factor (Lesmana., 2014). The influence of each factor caused the decline in the productivity of bored pile foundation work so that the value of the coefficient is used as the value of the model coefficient of each factor as shown in **Table 2.2**.

Table 2.2 Model Coefficient of Each Factor

Factor	Model Coefficient
Environtment	0,11
Equipment	0,89
Labor	0,51
Material	0,09
Management	0,77

Source; (Lesmana., 2014)

Table 2.3 Value Model Coefficient, Unit Expected Rate and Index Disruption

Factor	**Model Coefficient**	**Expected Unit Rate (m^1/hr)**	**Disruption Index**
ENVIRONTMENT	0,11	4,95	0,98
EQUIPMENT	0,89	4,17	0,82
LABOR	0,51	4,55	0,90
MATERIAL	0,09	4,97	0,98
MANAGEMENT	0,77	4,29	0,85

Source; (Lesmana., 2014)

After measuring model coefficient, the next step is to measure the amount of expected unit rate. This rate is expected productivity when a factor happened, which is the amount of baseline productivity reduced by model coefficient. In addition, disruption index is a value that describe the work hour efficiency when the disrupting factors occurred, which has the amount of expected unit rate divided by baseline productivity (Lesmana., 2014). In other word, the comparison between expected productivity and baseline productivity. Model coefficient value as well as expected unit rate and disruption index shown in **Table 2.3**.

2.5 Piling Process Productivity Factors

This one-level composed of a one-level hierarchical structure consists of 10 major qualitative factors, these 10 factors : operator efficiency, weather conditions, site conditions, job management, soil removal system, pouring system, mechanical problems, owner and/or consultant problems, site investigation, and productivity estimate accuracy. These productivity factors to provide a quantitative measurement for each factor's effect on productivity. In fact, these 10 factors have different attributes or categories that constitute the effect of each factor, concentrates only on the main factors without considering the sub factors or attributes.(Halpin, 2004)

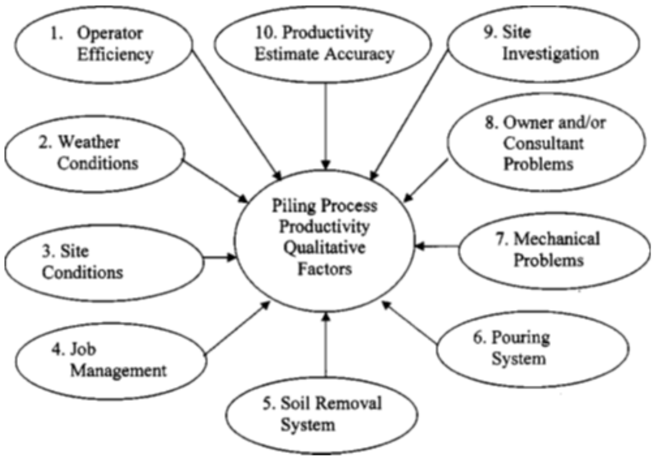


Figure 2.1 Piling Process Productivity Qualitative Factors
Source; (Halpin, 2004)

The factors that affect pilling construction have been identified by Peurifoy (1996): hardness of the soil; depth of holes; and time lost. The rate of pilling varies with the pressure of the air. Another item that influences the rate of pilling is the machine availability factor. Drills are subjected to severe vibration and wear, which may result in frequent failure of critical parts, or deterioration of the whole unit, entailing mechanical delays. The portion of time that a drill is operative is defined as the *availability factor*, which is usually expressed as a percent of the total pilling time.

Table 2.4 Identified Impact Factors Affecting Productivity

The Previous Research	Identified Impact Factors Affecting Productivity
Homyun Jang et al. (2011)	25 critical variables and were grouped into 4 groups, namely work management, work technique, work characteristic and worker component.
Durdyev and Mbachu (2011)	56 variables affecting construction productivity, and categorized them in to 8 factors of internal group and external group.
Enshassi (2007)	Classifies factors affecting productivity in the construction projects in to 10 groups, namely: factors associated with the internal workforce, factors associated with leadership, factors associated with work motivation. Factor associated with time, factors associated with materials and equipment, factors related to supervision, factors related to project characteristic, factors related to security, factors related to quality and external factors.
Soekiman et al. (2011)	113 variables affecting construction productivity and these variables were grouped into 15 groups of factors according to their characteristics, namely: Design, execution plan, material, equipment, labor, health and safety, supervision, working time, project factor, quality, financial, leadership and co-ordination,organization, owner/consultant and external factor.
Nabil Ailabouni et al. (2007)	32 significant variables affecting the productivity in the construction industry and these were grouped into 4 groups based on their characteristic, namely: environment factors, organizational factors, group dynamics and personal factors.

Source;(Durdyev, 2011; Enshassi., 2009; Gidado, 2009; Homyun Jang, 2011; Soekiman., 2011)

In summary, based on the previous research, 19 factors are selected and grouped into 5 groups according to their characteristics, namely: labor, management, environmental, material and equipment.

Efforts to produce better performance and increasing productivity in construction requires an understanding of the productivity various indicators as a path to understanding the performance of the project (Atkinson, 1997). It can essentially be done by reducing project cost overrun and project completion delay (PF Kaming, 1998).

Understanding critical factors affecting productivity of both positive and negative can be used to prepare a strategy to reduce inefficiencies and to improve the effectiveness of project performance. Knowledge and understanding of the various factors affecting pile construction productivity is needed to determine the focus of the necessary steps in increasing productivity and overall project performance.

Pile Construction productivity measurement is an analysis into the ratio of total output versus the total input of the construction process (Mawdesley M J, 2000). The measurement problem is exacerbated by the fact that the construction industry is composed of four sectors that differ significantly in; the outputs produced, firm size, and use of technology. The four sectors, which taken together define the construction industry, are residential, commercial/institutional, industrial, and infrastructure (Huang, 2009). This means that for each of these unique sectors, which in most cases construction companies are engaged in simultaneously, the contractor has to develop systems of measuring the different forms of output and the varied conditions under which all these projects are carried out.

Huang (2009) proposed what became the major delimitation criteria for this study when they identified that the construction industry has three distinct levels as far as productivity is concerned; task level, project level and industry level. *Task* refers to a specific activity like pouring concrete or structural steel erection, *projects* refers to the collection of activities which result in the renovation or the construction of a facility. *Industry* refers to the total portfolio of projects within an economy.

The construction company was an entity that had various projects where the actual productivity was taking place (Jonsson, 1996). This then places the projects in a key position with regards to productivity measurement as the ultimate products of construction. With this in mind this study will focus on project level productivity because construction productivity rates differ between projects because of the varying environments, characteristics, and project management efforts for each project (H.-S. Park, Thomas, S.R. & Tucker, R.L., 2005). Therefore, when analyzing construction productivity, one should consider the drivers that cause construction productivity differences between projects.

Paul W. Chan (2007) took the case for project based measurement of productivity further by highlighting that although strategic levels of management were crucial in improving construction productivity, there was a need to relate it to the projects themselves. They quote Groak (1994) who reinforces this by saying that industry had gone amiss by failing to recognise the project site as the "defining locus of production organisation". This meant that the industry needed to reorient its production improving efforts to the projects, and to do this measurement becomes critical.

Good project management in construction must vigorously pursue the efficient utilization of labor, material and equipment. Improvement of labor productivity should be a major and continual concern of those who are responsible for cost control of constructed facilities. Material handling, which includes procurement, inventory, shop fabrication and field servicing, requires special attention for cost reduction. The use of new equipment and innovative methods has made possible wholesale changes in construction technologies in recent decades. Organizations which do not recognize the impact of various innovations and have not adapted to changing environments have justifiably been forced out of the mainstream of construction activities. Job-site productivity is influenced by many factors which can be characterized either as labor characteristics, project work conditions or as non-productive activities.

Macro environmental forces refers to major external and uncontrollable factors that influence an organization’s decision making, and affect its performance and strategies. Six macro environmental forces are :

a. **Natural Environment**, the natural environment encompasses all living things and none living things occuring naturally on earth, such as changes scarcity of the raw material and increment weather with adverse to construction industry. The effect are increase price of sand, timber, rubble, bricks, and also make the portion of GDP decreases. (Industry, 2014)

b. **Political and Legal Environment**, the political environment refers to the actions taken by the government, which potentially affect the daily business activities of any business or company, such as change in legislature and executive, governmenr implement new plans and government bodies. The effect of this case, the government suspends work to review the project, incesement of government investments of roads, bridges, establish of National Adv. Council on Construction and Construction Industry Development Authority. (Industry, 2014)

c. **Economical Environment**, the economic environment consists of external factors in a business’s market and the broader economy that can influence a business, it consist with government, business firm, house hold, and foreign factor. The relative higher investment in economical infrastructural development are highways, airports, seaports, road development project.

d. **Demographical environment**, a demographic environment is a set of demographic factors such as gender or ethnicity. Increase population of women than the population of men. The effort is decrease the workforce of men.

e. **Cultural environment**, a cultural environment is a set of beliefs, practices, customs and behaviors that are found to be common to everyone that is living a certain population as a social infrastructure development. (Industry, 2014)

2.6 Pile Construction Productivity Prediction

Literature cites a number of productivity models that describe factors and predict productivity based on data collected. Contractor companies usually track construction productivity information and use their own historical productivity data to predict future projects. This predict can be used as a baseline for productivity and can be obtained by using historical data from similar projects (Sweis, 2000). The regression model is the most frequent statistical technique used to predict productivity (Sanders, 1993), (Smith, 1999), (H. R. a. Z. Thomas, I., 1999). This technique enables one to identify the impact of various factors and establish productivity predicts based on actual productivity data. Hanna et al. (Hanna, 1999a) used regression models to examine the impact of change orders on productivity for construction. Koehn and Brown (Koehn & 1985) established non-linear equations to explain weather effects on construction productivity. The learning curve also is an important factor in productivity. The learning curve theory states that the productivity of the same repetitive work will be continuously improved as a result of greater familiarity with the activity, better management, and more efficient use of tools and equipment (Oglesby, 1989). Mathematical learning curve models have been developed to predict productivity (H. R. Thomas, Mathews, C. T., and Ward, J. G., 1986); (H. R. a. Y. Thomas, I., 1987).

The productivity prediction model that considered the effects of project environment factors and management efforts. Raw productivity is defined as an observed value during construction. It includes the effects of project environment and management efforts performed by a project team. Therefore, raw productivity is not an absolute productivity value. For example, same workforce may produce different productivity rates in same work on different construction sites due to different environment even though their management efforts are same. It means one needs to detect the impacts of project environment and management efforts to calculate absolute productivity called by baseline productivity in the paper. Baseline productivity is pure productivity rates for each work. The previous study defined baseline productivity as the best productivity can be achieved and is unaffected by disruptions (H. R. a. Z. Thomas, I., 1999). Previous study considered only negative impact of project management in means of disruptions. Therefore their baseline productivity is the best performance. However, the rationale of this framework is that project management may also have positive impact on productivity. The concept of baseline in the paper is silmilar with adjusted productivity (H. R. a. Z. Thomas, I., 1999). It enables to estimate expected productivity during a project planning phase.

Prediction is a necessary assignment in construction management. It includes cost (bid preparation, budget), time (productivity, project schedule), or quality prediction. Despite from that, the prediction is complicated, intuitive and approximate. For the productivity prediction, there can be so many factors that influence the productivity of construction tasks because the tasks involve long sequential processes, craftsmanship, many materials and tools, and changeable site conditions. Some of the factors are easily recognized; some of them may not. Also, the extent of these factors affect the productivity is difficult to identify. To avoid these problems, SVR models have been successfully utilized in process productivity prediction.

2.7 Support Vector Machine

2.7.1 Overview of Support Vector Machine

Support Vector Machines (SVM) is a non-linear method that is based on machine learning. SVM was developed by Vapnik et al. based on Structural Risk Minimization principle from statistical learning theory (Joachims, 2002).

SVM is a technique for predicting, either classification or regression occurrence, machine learning has several techniques for in case classification, prediction, and regression. For classification case, if the data to be tested is not known label, it can use the cluster technique. The cluster technique has the option, among other hierarchical clustering or K-means. Prediction discrete data tables can be used to perform Linear Discriminant Analysis (LDA), Analytical Neural Network (ANN), Support Vector Machine (SVM), and others. It can also be used for regression case for continuous data output. The method can be used is linear regression, Support Vector Regression (SVR) or Analytical Neural Network (ANN).

SVR and ANN are feasible because of the higher power to rigorous procedures such as optimizing the parameters which are required for the proper use of these techniques, and can be implemented over a short timescale. The benefit of these two techniques is flexibility-meaning it is not restrained to be linear when presenting a calibration. The comparison between SVR and ANN is explained in **Table 2.5**:

Table 2.5 Comparison between Support Vector Regression and Artificial Neural Network

Support Vector Regression	1. To find a multivariate regression function referring to data with aim to predict an anticipated result property from a sample. 2. Effectual use to handling non-linear relationship in the data in easy method. 3. SVR being a more general model, shows higher robustness and better predictive ability for future samples. 4. SVR is able to overcome the overfitting

	5. The optimum parameter shown by the lowest root mean square error of cross validation (RMSECV) 6. SVR in modelling weighing error in the preparation at high significance level.
Artificial Neural Network	1. To find a relativity between inputs and outputs. A detailed input leads to a precise target output. The relation is adjusted referring to a comparison of the output and the target, until the network output practically equal to the target. 2. ANN is employed in signal processing, reduce and optimize the data, understanding and forecast of spectra and calibration. 3. ANN needs various controlling parameters, such as; number of hidden layers, number of neurons on each hidden layer, activation functions, number of training epochs, learning rate, and momentum term. To optimizing parameter of ANN, need to through some experiments which can recover the model performance. 4. ANN cannot be simply employed for generalization purposes. 5. Composed of three layers, it is introducing of input data (hidden layer), correcting and adjusting inputs by weight, and output layer (connection to the outside world). More hidden layers may cause overfitting. 6. The outputs are compared with targets and the difference between them is called error. 7. ANN is iterated several times till the error

Source : (Naguib, 2012; Wang, 2012)

Support Vector Machine (SVM) is essentially an input/output machine. A user is able to put in an input, and based on the model developed through training, it will return an output. It can then be applied to accurately predict the outcomes of upcoming cases (Shah, 2007). Data analysis and patterns recognize are applied in machine learning by SVM. SVM is in the same class as Artificial Neural Network (ANN) in terms of solving the functionality and condition of problems. In most implementations, SVM gives a better solution than ANN. ANN is only able to find a local solution, otherwise SVM can provide optimal global solution (Santosa, 2010). The characteristics of SVM consist of linear classifier, pattern recognition, Structural Risk Minimization (SRM), and it can handle two-class classification. Meanwhile, the advantages of SVM include: generalization capability, curve of dimensionality, feasibility, strong implication ability, fast learning capacity, capability for accurate predictions (Lam, Palaneeswaran, & Yu, 2009). SVM has occurred as a significant learning technique for resolving the classification and regression problems in varied subjects (Parrella, 2007). The concept of Support Vector Regression (SVR) is to maximize hyperplane to get data into a support vector. The one advantage of SVR is being able to overcome overfitting (Yasin, Prahutama, & Utami, 2014).

2.7.2 Support Vector Machine Classification

The objective of SVM Classification is to realize the best hyperplane as separation into two classes. This is obtained by calculating the boundary from a hyperplane and finding out the considerable boundary. The margin boundary is the length between hyperplane and the nearest data from every single class namely “Support Vector”.

Characteristics of SVM are: linear classification, pattern recognition which reveals transformed data in input space to high dimensional, optimized in the new vector, Structural Risk Minimization (SRM), and it can handle 2 class classification. Advantages of Support Vector Machine are generalization ability, cause of dimensionality, feasibility, strong inference capacity, fast learning capacity, and ability for accurate predictions. Corrected in a linear way. It works out that in most instances the optimization can be solved more easily in its binary interpretation. Moreover, the binary interpretation provides the solution for spreading sv machine to nonlinear functions. Therefore we will employ a paradigm dualization method utilizing LaGrange multipliers.

2.7.3 Support Vector Machine Regression

Support Vector Machines algorithms have been advanced for numeric prediction. One variation SVM employed for regression approach is Support Vector Regression (SVR). The idea of maximizing SVR is employing hyperplane to obtain data into a support vector. One of SVR’s advantages is being competent to overcome overfitting. Problem that is frequently experienced when employing the SVR includes determining the optimal parameters.

The fundamental idea of SVR is to plot data in a high-dimensional vector space by plotting non-linear and linear regression in the space. This technique can also be used directly into the regression, to preserve the distinctive of maximum margin algorithm. Nonlinear function is advanced using kernel functions that create the high-dimensional vector space in which the system is observed throughout the parameters that are not appropriate on the dimension of vector space. SVMs were first advanced for pattern recognition. When the SV algorithm was widespread to the instances of regression estimation, it was critical to discover a method of holding this feature. In pattern recognition, when measuring the damage occurred for a particular pattern, there is a great region where we accrue zero loss: every time a pattern is on the precise side of the judgment surface, and does not cover the margin, it does not participate any damages to the objective function. A damage function for regression estimation must have an insensitive area as well, therefore SVR employed the ε-insensitive loss (Scholkopf, 2001).

The prediction output could come out by cross-validation according to the “leave-one-out” assumption. K-fold cross validation suggest separating the data training set into k folds of equal sizes, in which each of the k folds is employed as a validation set and the other k-1 folds employed as training instances. Then, every instance is predicted once. The prediction outputs are averaged across the distinct folds to evaluate the performance. The rationale behind cross-validation involves removing the possibility of over fitting and strengthening the generalization ability of the regression.

Performance measure in this study used Root –mean Squared Error (RMSE) by this formula:

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{actual1} - y_{predict1}\right)^2} \qquad \mathbf{(4)}$$

Where n = the total number of data;

$y_{actual1}$ = the observed value;

$y_{predict1}$ = the predict value.

2.8 Time series

Time series is a compilation of data observation within a timeline. Time series method is a forecasting method using relationship pattern between variable which was estimated using time variable. Time series forecast differentiate based on data types and pattern. Generally, there are four kinds of time series data, which are horizontal, trend, seasonal, and cyclic (Subagyo, 1986). Horizontal pattern is a random, unexpected event, but its appearance will affect time series data fluctuation. Trend pattern is data tendency within long term, which may take form as increasing or decreasing pattern. Seasonal pattern is a data fluctuation that happens periodically within a year, such as quarterly, monthly, weekly, or daily. Cyclic pattern is a data fluctuation that happens longer than a year.

Box & Jenkins (Box, 1976) states that time series is an accumulation of observed data that were compiled within a timeline, and were used in order to see differences within such timeline. Time series data itself is data that was compiled, recorded, and observed in sequence. Observation period may range from annual, monthly, weekly, and in several cases, may be daily or even hourly. Time series data prediction is data assumption based on the said data's history. Its purpose is for finding a pattern within a historical data timeline and uses the found pattern to predict the future data. That data prediction may be used if it fulfills these three conditions: a.) historical data were provided, b.) Information can be quantified using numerical data, c.) It is assumed that pattern aspects will be repeated in the future (Makridakis, 1998).

2.9 R Program

R is a language and environment for statistical computing and graphics. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.

R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering) and graphical techniques, and is highly extensible. It is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.

One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control. R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes

1. an effective data handling and storage facility,
2. a suite of operators for calculations on arrays, in particular matrices,
3. a large, coherent, integrated collection of intermediate tools for data analysis,
4. graphical facilities for data analysis and display either on-screen or on hardcopy, and
5. a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

The term “environment” is intended to characterize it as a fully planned and coherent system, rather than an incremental accretion of very specific and inflexible tools, as is frequently the case with other data analysis software.

CHAPTER III: DATA COLLECTION AND ANALYSIS

3.1 Research Framework

The study mechanism describes that the research was based on the problems of inaccuracies of productivity of pile construction projects in Indonesia. It then identify the factors affecting pile construction productivity on projects in Indonesia. After acquiring influential factors, this data were collected from various projects in Indonesia. Using these datas, we can search for the baseline value of the pile construction productivity loss. The baseline is used as the evaluation of the accuracy to predict pile construction productivity. The last step of this part is factor analysis, resulting in macro impact factor for pile construction project productivity.

The second phase is data analysis. It begins by statistically analyzing and describing datas obtained from the first phase. The factors and datas were arranged in order to transform qualitative data into quantitative data. This step was carried out during data standardization in which all datas are expressed as a number between 0 if the factor does not occur in a working day and 1 if the factor occurred in a working day. After that, the data are converted into CSV file which was inserted to the data processor as a requirement for the model.

Third phase is developing model prediction by using support vector regression method. This phase starts by inputting transformed CSV file into R software and choosing syntax as classifier. The model development we using regression, specifically Support Vector Regression. The most important process at this stage is to set the parameters that will determine the accuracy of the model. The parameters consist of x and y parameters, in which y represents kernel. The validation test made use of Root Mean Square Error to checks into test accuracy (error rate) and correlation rate from the actual value to predicted value. If the accuracy value is greater than minimum value, the process is repeated from the 1st phase by reducing the factors. This last stage goes on until the maximum accuracy value is derived to verify the proposed model.

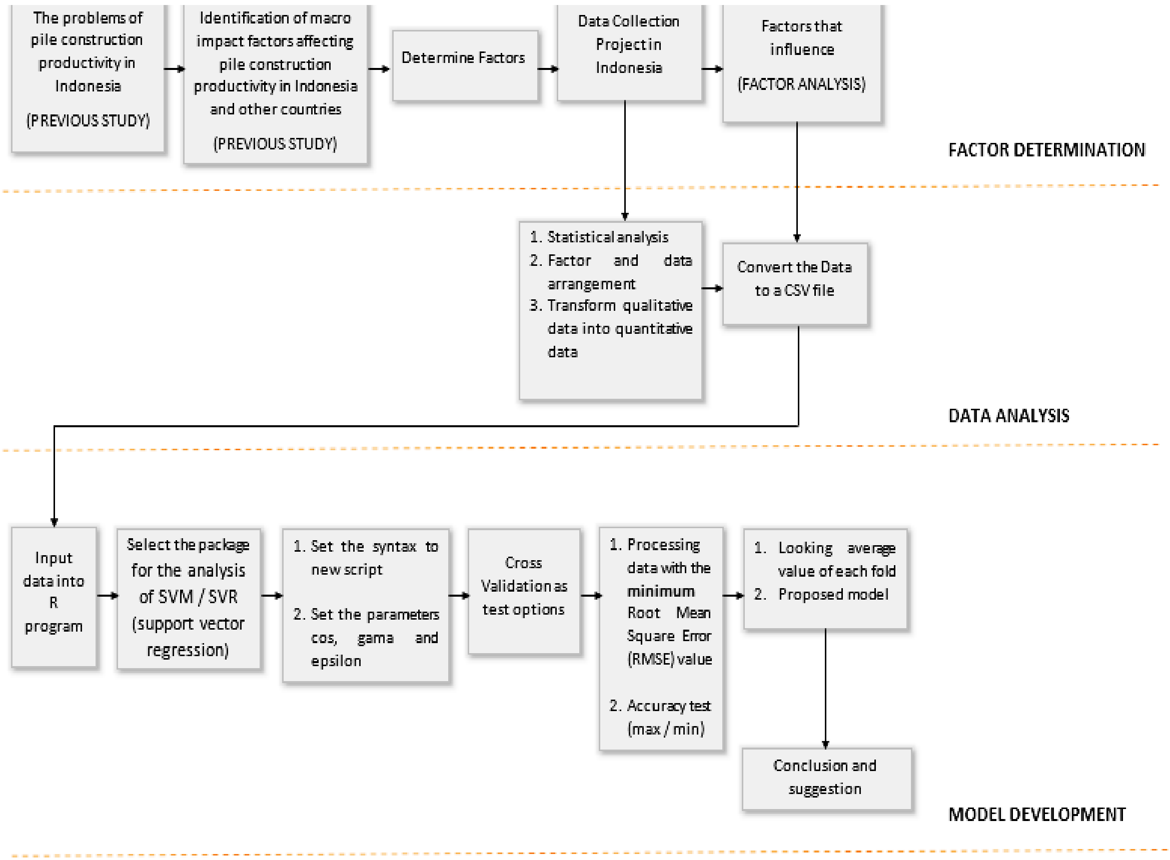


Figure 3.1 Research Procedure

3.2 Data Analysis

3.2.1 Overview Data

110 data used in the research were taken from completed construction projects, published from 2011-2016. All of the projects are building projects which have different types and characteristics are related to the purpose of this research. The comparison determines the macro impact factor pile construction productivity in Indonesia. The data should fulfill the following criterias:

Table 3.1 Location Characteristics in Indonesia

LOCATION Characteristic	**JAKARTA**	**SURABAYA**	**MALANG**	**SOLO**	**SEMARANG**
SOIL CONDITTION	Bad soil condition / peatland	Clay + sand (mud/silt soil)	Subgrade soil / rock soil	Clay	Clay + sand + gravel
TYPE OF PILE/foundation	Raft foundation	• Prestressed • Precast piles • Anchor foundation	Bored Pile (cast in situ)	• Prestressed / precast piles • *micropiling* or *helical piling*	• Prestressed/precast piles + bored pile • Anchor foundation

LOCATION Characteristic	JAKARTA	SURABAYA	MALANG	SOLO	SEMARANG
Size of the Project (average 864 m2 – 1548 m2)	• Grade 7 • Project of State Minister for State Owned Enterprises • > 50 billion				

3.2.2 Project Characteristics

This research used 110 project sample from state-owned project, with grade 7 building (highest grade) and the area of the project is somewhere between 865 m^2-1538 $m^{2.}$

Table 3.2 Construction Company Qualification

Qualification	Sub-Qualification	Asset	Grade
Small Company	K1	The ability to do job valued 1 billion rupiah	Grade 1 – 4
	K2	The ability to do job valued 200-350 million rupiah	
	K3	The ability to do job valued 350 -500 million rupiah	
Medium Company	M1	The ability to do job valued 833 million up to 2 billion rupiah	Grade 5
	M2	The ability to do job valued 2-10 billion rupiah	
Large Company	B1	The ability to do job valued 10-50 billion rupiah	Grade 6 – 7
	B2	The ability to do job valued 50 -83,3 billion rupiah	

Table 3.3 Sample Data Collected

No	Date	Pile Depth	Number of Pile Finish/day	Total Dept	Work Hour/day	Equipment Sheet	Total Work Hour of Equipment	Daily Productivity	Factor				
		m	qty	m	hour	qty	hour	m / hour	Env.	Equip.	Labor	Material	Manage
		1	2	3 = 1 x 2	4	5	6 = 4 x 5	7 = 3 / (4*5)					
1	10/11/2014	13	6	78	12	4	48	1.63	0	0	1	0	0
2	10/12/2014	12	5	60	14	3	42	1.43	1	0	0	1	1
3	10/13/2014	14	5	70	13	3	39	1.79	0	0	0	0	0
4	10/14/2014	12	6	72	15	2	30	2.40	0	0	1	0	0
5	10/15/2014	14	4	56	10	2	20	2.80	1	0	0	1	0
6	10/16/2014	14	3	42	10	2	20	2.10	0	0	0	0	0
7	10/17/2014	14	5	70	11	4	44	1.59	0	1	0	0	0
8	10/18/2014	13	6	78	15	3	45	1.73	0	0	1	0	0
9	10/19/2014	12	8	96	18	3	54	1.78	0	0	0	0	0
10	10/20/2014	12	6	72	15	3	45	1.60	1	0	1	0	0
11	10/21/2014	13	4	52	18	3	54	0.96	0	0	0	0	0
12	10/22/2014	11	4	44	12	3	36	1.22	0	0	1	0	0
13	10/23/2014	11	3	33	10	3	30	1.10	0	0	0	1	0
14	10/24/2014	10	3	30	14	2	28	1.07	0	1	0	0	0
15	10/25/2014	13	3	39	11	3	33	1.18	0	0	1	0	0
16	10/26/2014	14	4	56	15	3	45	1.24	1	0	0	0	0
17	10/27/2014	14	4	56	16	4	64	0.88	0	0	1	0	0
18	10/28/2014	15	4	60	11	4	44	1.36	0		0	1	0

No	Date	Pile Depth	Number of Pile Finish/day	Total Dept	Work Hour/day	Equipment Sheet	Total Work Hour of Equipment	Daily Productivity	Factor				
		m	qty	m	hour	qty	hour	m / hour	Env.	Equip.	Labor	Material	Manage
		1	2	3 = 1 x 2	4	5	6 = 4 x 5	7 = 3 / (4*5)					
19	10/29/2014	12	3	36	15	4	60	0.60	0	1	0	0	0
20	10/30/2014	13	2	26	13	2	26	1.00	0	0	1	0	1
21	10/31/2014	12	4	48	15	3	45	1.07	1	0	0	0	0
22	11/1/2014	14	3	42	18	3	54	0.78	0	0	0	0	0
23	11/2/2014	13	3	39	15	3	45	0.87	0	0	1	0	0
24	11/3/2014	12	3	36	10	3	30	1.20	0	0	0	1	0
25	11/4/2014	14	4	56	14	2	28	2.00	0	1	0	0	0
26	11/5/2014	13	3	39	16	3	48	0.81	0	0	1	0	0
27	11/6/2014	12	4	48	15	3	45	1.07	1	0	0	0	0
28	11/7/2014	15	6	90	17	4	68	1.32	0	1	0	0	0
29	11/8/2014	12	4	48	16	4	64	0.75	0	0	0	1	0
30	11/9/2014	14	5	70	15	4	60	1.17	0	0	1	0	0
31	11/10/2014	12	2	24	10	2	20	1.20	0	0	1	0	1
32	11/11/2014	12	3	36	10	3	30	1.20	1	0	0	0	1
33	11/12/2014	12	7	84	16	3	48	1.75	0	0	1	0	0
SUMMARY			**139**				**1,392**		**7**	**5**	**13**	**6**	**4**

Data on the table above is a sample from a project occurred on 2014 in Java. Based on the factor table, number 0 represents obstacle that occurred onsite during the mentioned dates, while 0 means no obstacle occurred.

Location characteristic is an important factor that need to be considered in order to predict the pile construction productivity. Conditions, environmental aspect, and site project location are sub-factors that are included in location characteristic.

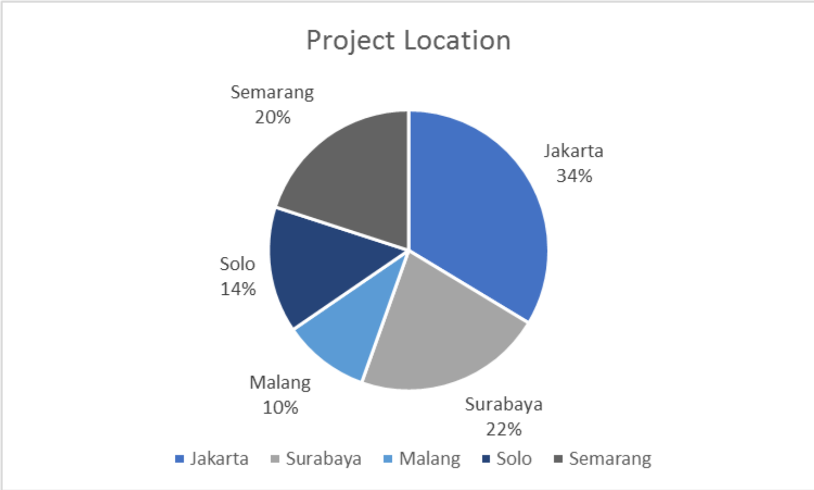


Figure 3.2 Project Location Chart

The type of pile used on construction projects in Indonesia are generally conventional materials, i.e. concrete and steel. For concrete material, constituent materials of concrete have small fluctuation. Each type of foundation gives different effect to project productivity because each type of foundation has a different characteristic. The foundation type on each respective cities in Indonesia are described in **Figure 3.3**.

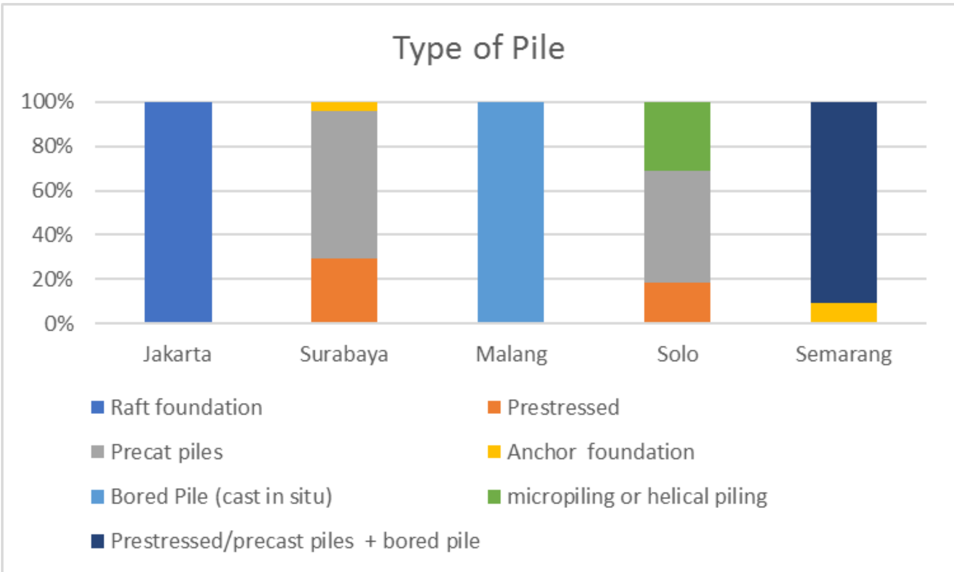


Figure 3.3 Type of Pile Chart

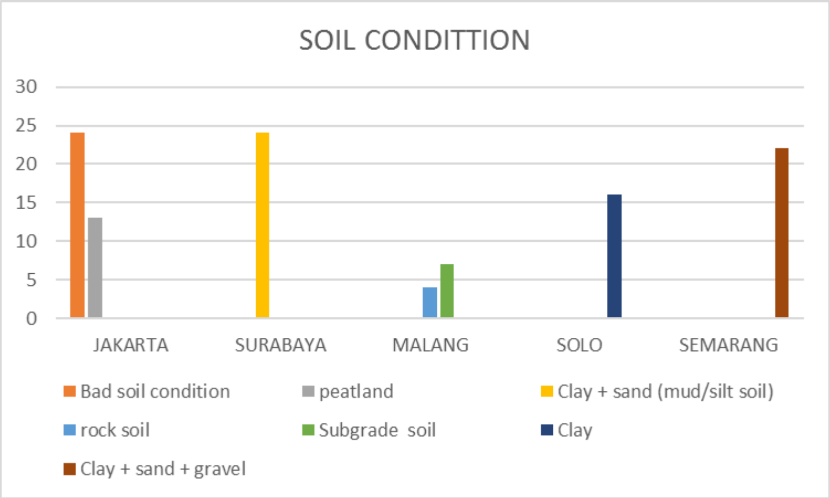


Figure 3.4 Soil Condition Chart

In summary, based on the previous research and data, 19 factors are selected and grouped into 5 groups according to their characteristics, namely: labour, management, environmental, material and equipment as shown in the **Table 3.4.**

Table 3.4 Impact Factors Affecting Productivity

Group	Factors
Labour	Lack of labour skills Increase of labourer age Labour absenteeism Lack of training Labour personal problem
Management	Poor site management Poor communication Misunderstanding between labour and supervisor Lack of periodic meeting with labors Quality inspection delay
Environmental	Weather changes Project location Working with confined place Large project size
Material	Material shortage Unsuitable material storage location Low quality raw materials
Equipment	Old and inefficient equipment Tools and equipment shortages

3.3 Factor Analysis

Factor analysis is a necessary step to improve the model efficiency. This step will determine the rank of influential factor. This research uses factor extraction as analysis method. Factor extraction involves determining the smallest number of factors that can be used to best represent the interrelations among the set of variables. There are a variety of approaches that can be used to identify (extract) the number of underlying factors or dimensions. Some of the most commonly available extraction techniques. This involves balancing two conflicting needs: the need to find a simple solution with as few factors as possible; and the need to explain as much of the variance in the original data set as possible.

The insignificant factor will listed and ranked to determine which factors are less likely influential in accuracy of project pile construction productivity. The data were obtained from construction projects which are spread throughout Indonesia. Factors that were collected were analyzed using factor analysis using SPSS software. This step begins by inputting the data obtained from the project.

Table 3.5 Macro Impact Factors Input Data Sample

Project Data	**Impact Factor**					**Pile Construction Productivity**	**Total Factor**
	Env	**Equip**	**Labor**	**Material**	**Mangement**		
1	7	10	14	5	11	26	47
2	6	9	10	6	9	21	40
3	5	8	8	7	8	19	36
4	7	7	9	7	10	20	40
5	14	16	20	15	18	41	83
6	16	8	20	19	12	30	75
7	16	6	20	17	12	28	71
8	16	6	20	17	12	28	71
9	16	8	23	20	11	31	78
10	11	10	19	14	7	26	61
11	11	10	18	13	7	26	59
12	12	6	11	16	10	21	55

Table 3.5 shows data sample that was inserted in order to calculate the correlation between 12 out of 110 factors using SPSS program.

Asumption 1: Relationships Between Variables is Good

Correlations

	Environment	Equipment	Labor	Material	Management

Environment	Pearson Correlation	1	.382**	.807**	.896**	.556**
	Sig. (1-tailed)		.000	.000	.000	.000
	N	110	110	110	110	110
Equipment	Pearson Correlation	.382**	1	.467**	.289**	.624**
	Sig. (1-tailed)	.000		.000	.001	.000
	N	110	110	110	110	110
Labor	Pearson Correlation	.807**	.467**	1	.779**	.555**
	Sig. (1-tailed)	.000	.000		.000	.000
	N	110	110	110	110	110
Material	Pearson Correlation	.896**	.289**	.779**	1	.483**
	Sig. (1-tailed)	.000	.001	.000		.000
	N	110	110	110	110	110
Management	Pearson Correlation	.556**	.624**	.555**	.483**	1
	Sig. (1-tailed)	.000	.000	.000	.000	
	N	110	110	110	110	110

**. Correlation is significant at the 0.01 level (1-tailed).

Correlations

			Environment	Equipment	Labor	Material	Management
Kendall's tau_b	Environment	Correlation Coefficient	1.000	.285**	.626**	.754**	.369**
		Sig. (1-tailed)	.	.000	.000	.000	.000
		N	110	110	110	110	110
	Equipment	Correlation Coefficient	.285**	1.000	.320**	.190**	.048
		Sig. (1-tailed)	.000	.	.000	.004	.262
		N	110	110	110	110	110
	Labor	Correlation Coefficient	.626**	.320**	1.000	.630**	.141*
		Sig. (1-tailed)	.000	.000	.	.000	.026
		N	110	110	110	110	110
	Material	Correlation Coefficient	.754**	.190**	.630**	1.000	.242**
		Sig. (1-tailed)	.000	.004	.000	.	.000
		N	110	110	110	110	110
	Management	Correlation Coefficient	.369**	.048	.141*	.242**	1.000
		Sig. (1-tailed)	.000	.262	.026	.000	.
		N	110	110	110	110	110
Spearman's rho	Environment	Correlation Coefficient	1.000	.362**	.771**	.882**	.488**
		Sig. (1-tailed)	.	.000	.000	.000	.000
		N	110	110	110	110	110
	Equipment	Correlation Coefficient	.362**	1.000	.393**	.228**	.078

		Sig. (1-tailed)	.000	.	.000	.008	.208
		N	110	110	110	110	110
	Labor	Correlation Coefficient	.771**	.393**	1.000	.762**	.254**
		Sig. (1-tailed)	.000	.000	.	.000	.004
		N	110	110	110	110	110
	Material	Correlation Coefficient	.882**	.228**	.762**	1.000	.402**
		Sig. (1-tailed)	.000	.008	.000	.	.000
		N	110	110	110	110	110
	Management	Correlation Coefficient	.488**	.078	.254**	.402**	1.000
		Sig. (1-tailed)	.000	.208	.004	.000	.
		N	110	110	110	110	110

**. Correlation is significant at the 0.01 level (1-tailed). *. Correlation is significant at the 0.05 level (1-tailed).

Asumption 2: KMO must be > 0,5 (satisfied)

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		**.780**
Bartlett's Test of Sphericity	Approx. Chi-Square	397.197
	df	10
	Sig.	.000

Some projects that have same function and characteristics might use different foundation type, depending on the soil condition of the site. Building coverage ratio does not represents the specification for the whole projects because its ground floor area as the standardized, while the floors on each project are different. Location supposed to have larger effect in prediction, but in this case, almost all of the location are located in accessible urban area, resulting location as an insignificant factor because Indonesian contractor will most likely choose concrete as structure material for all kind of building.

Macro indicators form an important part of the total productivity framework, but information regarding performance at the company level is of greater interest to most managers. This information comes under micro measurements. Some form of overall measurement will be available within each project. In terms of output, individual project contributes to the total output of the company, thus the outputs of all similar companies will make up the total industry. The outputs from the various industries, when aggregated, yield the national output. Hence, productivity is a concept which is equally relevant to all levels of the economy, be it individual, company, industry, or nation. Identifying major factors influencing productivity in an objective manner will result in establishing cause and effect relationships between individual factors and the calculated productivity.

It is important to differentiate between macro-and micro-level impact factors for pile construction productivity in order to analyze cause/effect relationships and take appropriate action. Macro-level factors that influence the effectiveness of construction are those that often attract considerable rhetoric, but not enough specific actions or economic support. There is abundant scope that can be done by both the industry as well as government to enhance and promote an atmosphere for sustained progress.

Government, industry, and the financial community must eventually cooperate to provide synergistic support at a macro-level. The industries that reinvest sufficient resources to remain competitive will survive and therefore the industry and construction companies must continue, and indeed increase, their contributions at a macro-level.

These are the macro-level concerns that needs to be addressed. However, the individual construction company or person has an obligation to improve productivity at a micro-level. Labor and methods efficiency are the foundation for competitiveness and more effort must be paid to productivity measurement.

It may not be possible to separate the individual effects of all influences. However, this should not deter efforts to quantify their effects and impacts on construction efficiency. Productivity improvement and effect measurement of related factors must become part of the daily construction routine. Both the contractor and the end user benefited from the improvement of construction quality and productivity. The production function produces data, which are analyzed and provide feedback for action toward improvement. The cycle continues until the required level of productivity is reached.

An analysis of productivity is complete only when quality and safety are also considered. The construction industry must become serious if it is to improve in quality, safety, and productivity. Supervisors and tradesmen must continue to improve their efficiency. Improvement is a continuous process, and an integral part of total quality management.

3.4 Macro Impact Factors For Pile Construction Productivity

Loss of productivity is the value of the model coefficient multiplied by the binary number of factors that occurred on each day of the study average from the first day until the 40th day, while work hours lost an hour's worth of lost work due to factors occurring on that day.

$$\mathbf{\Delta}\ \textbf{(deviation)} = \boldsymbol{Average - Act\ Productivity} = 0{,}11\,X_1 + 0{,}89\,X_2 + 0{,}51\,X_3 + 0{,}09\,X_4 + 0{,}77\,X_5 \qquad \textbf{(5)}$$

Table 3.6 Predictions Sample Model Coefficient of Each factors

Total Work Hours (hour)	Factor Affecting Pile Productivity					Loss of Productivity (m/ hour)				
	Env.	Equip.	Labor	Material	Manage	Env . (0.11)	Equip. (0.89)	Labor (0.51)	Material (0.09)	Manage (0.77)
32	0	0	0	0	1	0	0	0	0	0.77
22	1	0	1	0	0	0.11	0	0.51	0	0
0	0	0	0	0	0	0	0	0	0	0
68	0	0	1	1	0	0	0	0.51	0.09	0
42	0	1	1	0	0	0	0.89	0.51	0	0
45	0	1	0	0	0	0	0.89	0	0	0
48	1	1	0	0	0	0.11	0.89	0	0	0
8	0	0	0	0	1	0	0	0	0	0.77
SUMMARY										
265	**2**	**3**	**3**	**1**	**2**	**0.22**	**2.67**	**1.53**	**0.09**	**1.54**

Table 3.7 Predictions Findings Sample

Workhours Lost (hour)				
Environment	**Equipment**	**Labor**	**Material**	**Management**
0	0	0	0	4.47
0.44	0	2.04	0	0
0	0	0	0	0
0	0	6.30	1.11	0
0	6.79	3.89	0	0
0	7.27	0	0	0
0.96	7.76	0	0	0
0	0	0	0	1.12
SUMMARY				

1.40	**21.82**	**12.23**	**1.11**	**5.59**

After measuring model coefficient using SPSS program based on the previous research, the next step is to measure the amount of productivity loss. This table shows expected loss productivity when a factor happened, which is the amount of loss productivity multiplied by model coefficient. In addition, work is a value that describe the work hour efficiency when the disrupting factors occurred, which has the amount of loss productivity multiply by total work hours. In other word, the comparison between lost productivity and lost work hours. Model coefficient value compared to the findings (predictions) as well as loss productivity and work hours lost shown in **Table 2.2** and in **Table 2.3**.

The value of work hours lost is equal to loss of productivity multiplied by the working hours is then divided by baseline productivity. The calculations for the loss of value of lost productivity and work hours from 110 project shown in **Table 3.8** and the percentage shown in **Figure 3.5**. **Table 3.8** shows the details for lost working hours:

a. Factors equipment raises work hours lost by 7828,07
b. The factors causing labor work hours lost by 2422,75
c. Factors causing work hours lost management by 1574,14

Table 3.8 Lost Working Hour from 110 Project Caused by Macro Impact Factors

Factor	Lost Working Hours
Equipment	7.828,07
Labor	2.422,75
Management	1.574,14
Total Work Lost Hour	11.824,90
Total Workday (Hour)	175.601

The lost working hours were 175.601 hours out of 11.824,90. The factor that reduced productivity the most was equipment because on several working days, there might be some damaged equipment thus causing idle time that greatly reduce project productivity. The other factor is labor, because they often took a rest outside the break time, thus also causing idle time that decreases productivity. Researchers recommend the contractor to create a stricter rule with clearer penalty to minimize idle time and work hours lost, thus improving productivity and resulting in on-time project completion.

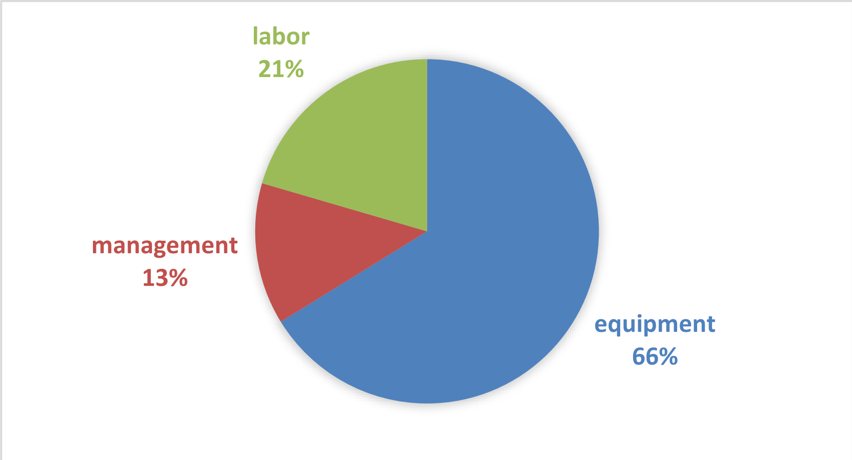


Figure 3.5 Macro Impact Factor (Cause of Productivity Loss)

From **Figure 3.5**, Equipment is a major part of 110 construction projects. In the process, Equipment is often regarded as somewhat simple, but in reality, the installation of equipment takes more time. All equipment metrics are calculated in hours/each, which is included in pile productivity. The size of equipment may make a difference on the productivity rate, however, the total weight or capacity of the equipment is surveyed to establish the relationship between the equipment and productivity.

CHAPTER IV: MODEL DEVELOPMENT

R is an open-source environment for statistical computing and visualization. It is based on the S language developed at Bell Laboratories, and is the product of an active movement among statisticians for a powerful, programmable, portable, and open computing environment. It is applicable even for the most complex and sophisticated problems, as well as "routine" analysis, without any restriction on access or use. The aim of this paper is to illustrate the capability of the R language.

In order to do the illustration, excel data is converted to a CSV file and then inputted into the program using the syntax for calling the data and the SVM can be run (SVR part of SVM, which is a tool to predict the parameter of SVR). From the modeling results, it is known that RMSE values are the average value of this error and correlation. A good model can be obtained if it has high correlation value and small RMSE value. It can be seen from the plot that there is a high productivity value that is completely different from the other. However, the correlation is still high because the prediction result follows a pattern. In that model, there are three parameters for SVR those are *cos, gamma, epsilon*. Each project data can be predicted by using y formula. In the graph *y*-axis = the value of productivity loss, whereas *x* is the time of the project. Overall results of the prediction actually show high correlation value, therefore it's suitable to forecast the next period. For the future, existing parameters can be included in the programs, and included factors that affects prediction results.

4.1 Arranging the Datasets

Using suitable random selection basis, the normalized datasets for a specific modelling constructs were arranged. Standardization is needed within the revised data because the data is divided into nominal factors and quantitative factors. The nominal factors such as type of location are entered into the urban as 0 and 1 for a sub – urban. The other nominal factors are describe in **Table 4.1**.

Table 4.1 Inputs and the coresponding labels for the model

Attributes	Label	Explanation	
Type of soil condition	0	Bad Soil Condition	These labels using **normalized number**, divided by total label each attributes, and converted to nominal within a factor of nominal 0 to 1
	0.14	Peatland	
	0.29	Clay + Sand (mud/silt soil)	
	0.43	Rock Soil	
	0.57	Subgrade Soil	
	0.71	Clay	
	0.86	Clay + Sand + Gravel	
Type of Pile	0.00	Raft foundation	
	0.17	Prestressed	
	0.33	Precat piles	
	0.50	Anchor foundation	
	0.67	Bored Pile (cast in situ)	
	0.83	Micropiling or Helical Piling	
Location	0	Urban	
	0.5	Sub-Urban	
City Location	0	Jakarta	
	0.2	Surabaya	
	0.4	Malang	
	0.6	Solo	
	0.8	Semarang	
Size of the project	These attributes use **normalized number**, then converted to nominal within a factor of nominal 0 to 1		
Pile depth			
Number of pile finish			
Number of equipment			

The attributes use normalized number, such as size of the project, pile depth, number of pile finish, and number of equipment, are converted to nominal within a factor of nominal 0 to 1. The equation for standardization/normalization is:

$$Norm = \frac{(Xdata - \text{data min})\ x(1-0)}{(data\ max - data\ \text{min}) + 0} \quad \textbf{(6)}$$

The examples of normalized into nominal number 0 to 1 are provide in **Table 4.2**.

Table 4.2 The normalized sample of projects from 4 Factors

Pile Depth	Number of pile	Size of project	Number of equipment	TRANSFORMATION DATA	Pile Depth	Number of pile	Size of project	Number of equipment
26	5	130	4		0.83	0.25	0.43	1.00
26	6	156	3		0.83	0.50	0.67	0.50
26	5	130	3		0.83	0.25	0.43	0.50
24	5	120	2		0.50	0.25	0.33	0.00
24	6	144	2		0.50	0.50	0.56	0.00
24	6	144	2		0.50	0.50	0.56	0.00
23	5	115	4		0.33	0.25	0.29	1.00
24	5	120	3		0.50	0.25	0.33	0.50
24	6	144	3		0.50	0.50	0.56	0.50
26	5	130	3		0.83	0.25	0.43	0.50
26	5	130	3		0.83	0.25	0.43	0.50
27	5	135	3		1.00	0.25	0.47	0.50
27	6	162	3		1.00	0.50	0.72	0.50
22	4	88	2		0.17	0.00	0.04	0.00
24	5	120	3		0.50	0.25	0.33	0.50
23	4	92	3		0.33	0.00	0.07	0.50
22	5	110	4		0.17	0.25	0.24	1.00
22	4	88	4		0.17	0.00	0.04	1.00
25	5	125	4		0.67	0.25	0.38	1.00
25	5	125	2		0.67	0.25	0.38	0.00
21	4	84	3		0.00	0.00	0.00	0.50
21	6	126	3		0.00	0.50	0.39	0.50
22	5	110	3		0.17	0.25	0.24	0.50
24	5	120	4		0.50	0.25	0.33	1.00
24	8	192	4		0.50	1.00	1.00	1.00
25	5	125	4		0.67	0.25	0.38	1.00
23	5	115	3		0.33	0.25	0.29	0.50
24	4	96	3		0.50	0.00	0.11	0.50
22	5	110	3		0.17	0.25	0.24	0.50
21	5	105	3		0.00	0.25	0.19	0.50

23	4	92	4		0.33	0.00	0.07	1.00
24	6	144	4		0.50	0.50	0.56	1.00
25	5	125	4		0.67	0.25	0.38	1.00
26	5	130	3		0.83	0.25	0.43	0.50
24	5	120	3		0.50	0.25	0.33	0.50

In order to input the data in SVR model, a data normalization that involves 8 factors is required because the unstandardized data consists of qualitative and quantitative data. Data normalization process can be seen in Table 4.3 and 4.4, where **Table 4.3** contains data before normalization and **Table 4.4** contains data after normalization.

Table 4.3 Sample Data Input Before Normalization

Project no.	Pile Depth	Number Of pile	Size of Project	Number of Equipment	Location	Soil condition	Type of pile	City
52	13	139	1786	3.03	1	0	0	0
102	23	223	5163	3.14	0	3	3	1
99	21	246	5052	3.10	0	1	1	2
98	21	244	5035	3.10	0	1	1	2
90	21	266	5428	3.10	0	1	1	2
70	23	227	5281	3.06	0	3	3	3
69	21	262	5352	3.10	0	1	1	2
64	22	215	4672	3.06	0	2	2	3
59	21	229	5021	3.03	0	1	1	0
58	22	214	4651	3.06	0	2	2	3
57	24	237	5516	3.14	0	4	4	1
38	21	230	4713	3.10	0	1	1	2
36	24	232	5401	3.14	0	4	4	1
26	32	210	6651	3.06	0	5	5	3
22	21	260	5357	3.10	0	1	1	2
16	24	213	5112	3.29	0	4	4	4
14	22	206	4480	3.06	0	2	2	3
13	22	223	4902	2.95	0	2	2	3
12	21	236	4843	3.10	0	1	1	2

11	13	234	3000	3.14	1	0	0	1
10	33	224	7376	3.14	0	6	0	1

Table 4.4 Sample Data Input After Normalization

Project no.	Pile Depth	Number Of pile	Size of Project	Number of Equipment	Location	Soil condition	Type of pile	City
52	0.00	0.52	0.23	0.08	0.50	0.00	0.00	0.00
102	0.50	0.84	0.70	0.18	0.00	0.43	0.50	0.20
99	0.38	0.92	0.68	0.14	0.00	0.14	0.17	0.40
98	0.38	0.92	0.68	0.14	0.00	0.14	0.17	0.40
90	0.38	1.00	0.73	0.14	0.00	0.14	0.17	0.40
70	0.50	0.85	0.71	0.10	0.00	0.43	0.50	0.60
69	0.38	0.98	0.72	0.14	0.00	0.14	0.17	0.40
64	0.44	0.81	0.63	0.10	0.00	0.29	0.33	0.60
59	0.38	0.86	0.68	0.08	0.00	0.14	0.17	0.00
58	0.44	0.80	0.63	0.10	0.00	0.29	0.33	0.60
57	0.53	0.89	0.74	0.18	0.00	0.57	0.67	0.20
38	0.38	0.86	0.63	0.14	0.00	0.14	0.17	0.40
36	0.53	0.87	0.73	0.18	0.00	0.57	0.67	0.20
26	0.93	0.79	0.90	0.10	0.00	0.71	0.83	0.60
22	0.38	0.98	0.72	0.14	0.00	0.14	0.17	0.40
16	0.53	0.80	0.69	0.32	0.00	0.57	0.67	0.80
14	0.44	0.77	0.60	0.10	0.00	0.29	0.33	0.60
13	0.44	0.84	0.66	0.00	0.00	0.29	0.33	0.60
12	0.38	0.89	0.65	0.14	0.00	0.14	0.17	0.40
11	0.00	0.88	0.40	0.18	0.50	0.00	0.00	0.20
10	1.00	0.84	1.00	0.18	0.00	0.86	0.00	0.20

After converting the data, it can be inserted to SVR model using 10 fold cross validation model in order to find the best correlation between data classes after classifying the data.

Table 4.5 Comparation Result between SVR Model with Normalization and without Normalization

No.				**Parameters**	**K = 10 fold**

	Total Mean Squared Error	**Squared Correlation Coefficient**	**Number of Support Vectors**		**0**	**1**	**2**	**3**	**4**
1	57.24	0.875	22	Cos = 1 Gamma = 0.125 Epsilon = 0.1	7.98	6.57	26.21	18.19	29.78
					32.12	15.96	15.56	37.12	14.09
2	**0.011**	**0.872**	**22**		**0.01**	**0.03**	**0.00**	**0.00**	**0.00**
					0.00	**0.03**	**0.02**	**0.00**	**0.00**

From the **Table 4.5**, the first calculation are the result from the training and testing data without normalization, it has big Mean Square Error = 57.24 but still has good Squared Correlation Value = 0.875, and the second one are the result from the training and testing data after normalization. It has good result, low Mean Square Error = **0.011** and also has high Squared Correlation Value = **0.872.**

The model explores the significant relationship between observed attributes (factor affecting) and predictor attributes. The whole data that consists of 8 attributes from 110 projects and is validated by cross validation method later on. The cross-validation method is quite popular for evaluation. This method divides the data into two parts, namely the training data and test data. To divide the data into several pieces, a method called 'Kfold', often called the 'Kfold Cross Validation', is used. The most popular division is to divide the data into 10 sections, known as 10-way fold.

In cross-validation approach, each record will be used several times within the same amount for training and testing. To illustrate this method, it is assumed that the data were partitioned into two subsets with the same size. One of the data will be used for subset, while the other is for testing purpose, and for the second round, the first data will be used for testing meanwhile the others will be used as subset. This approach is called two-fold cross-validation. Total error obtained using the accumulated errors from the two processes.

The second method was to use one data set for training and the other for testing. K-fold cross validation method generalized this approach by segmenting the data into k partition with the same size. During the process, one of the partition were used for testing, while the rest were used for training. This procedure was repeated for k times until all the partition were used for testing at least once. Error accumulation were obtained by the summation of all procedure's error. (Han, 2006).

K-fold cross validation method define k = N as the size of data set. This method is called as leave-one-out method, each test set should only contain one record. This approach has the benefit of using as much as data as possible for training. Test set tends to be mutually exclusive and effectively covers all the data set. The downside of this data is the amount of computation needed to repeat the procedure for N-times. K-fold cross validation is a technique to evaluate model accuracy, with these characteristics (Nugroho, 2008) :

a. Partition the data randomly into k fold, those are D_1, D_2,… Dn. Each data set has almost the same amount of data.
b. In repetition, i is used by Di as a testing data and other set as training data. For example:
    - In the first testing, D_1 acts as testing data while D_2… D_n acts as training data
    - In the first testing, D_2 acts as testing data while D_1, D_3… D_n acts as training data, etc.
c. Has training and testing time for k times
d. The accuracy was tested using a formula.

Therefore at the initial stage, data is divided into 10 parts randomly **(Figure 4.1).** Furthermore, different combination between the test data and the training data was performed to achieve the optimal parameter and the maximum accuracy. Training and testing was carried out where one data serves as an examiner and other data as training data.

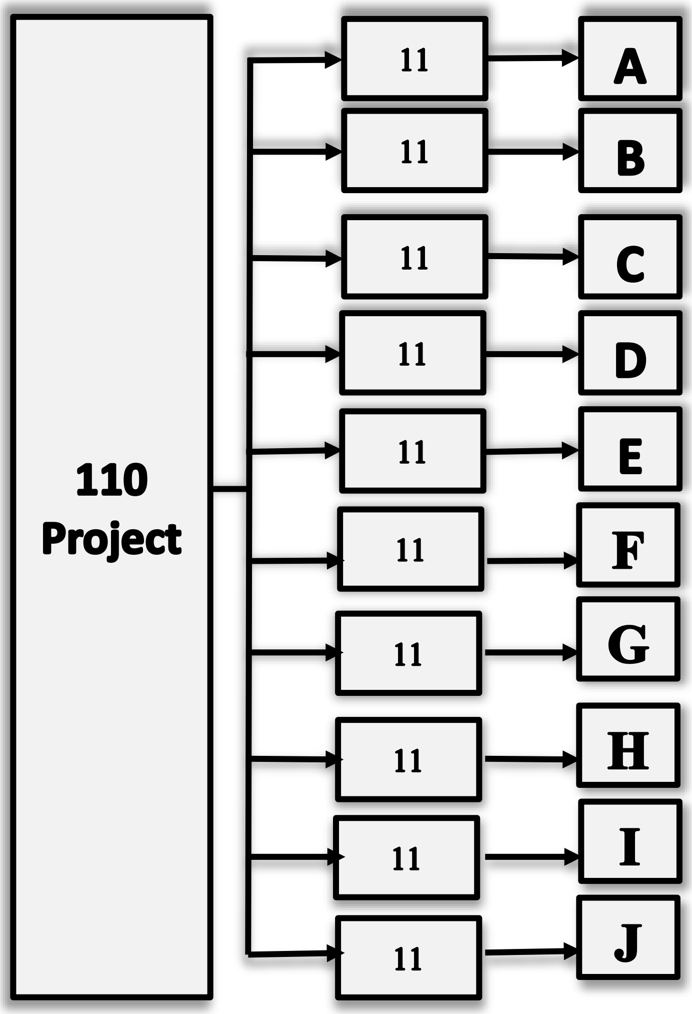


Figure 4.1 Folds Illustration

Table 4.6 Combination of 10 Folds Cross Validation

Combination	Training Data	Test Data
1	B, C, D, E, F, G, H, I, J	A
2	A, C, D, E, F, G, H, I, J	B
3	A, B, D, E, F, G, H, I, J	C
4	A, B, C, E, F, G, H, I, J	D
5	A, B, C, D, F, G, H, I, J	E
6	A, B, C, D, E, G, H, I, J	F
7	A, B, C, D, E, F, H, I, J	G
8	A, B, C, D, E, F, G, I, J	H
9	A, B, C, D, E, F, G, H, J	I
10	A, B, C, D, E, F, G, H, I	J

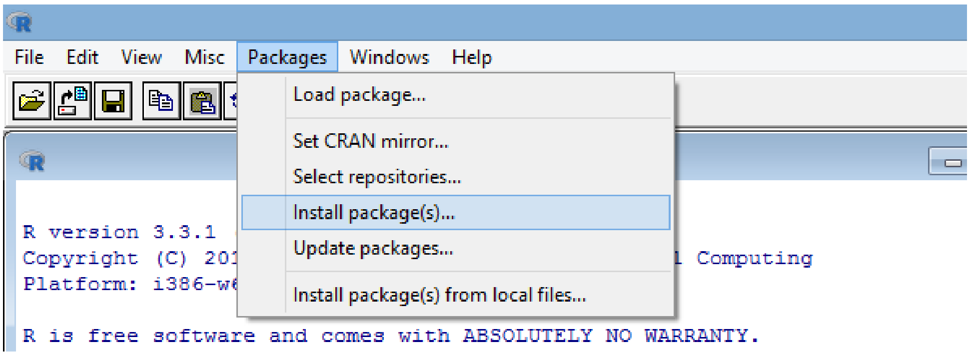
4.2 Model Implementation and Evaluation

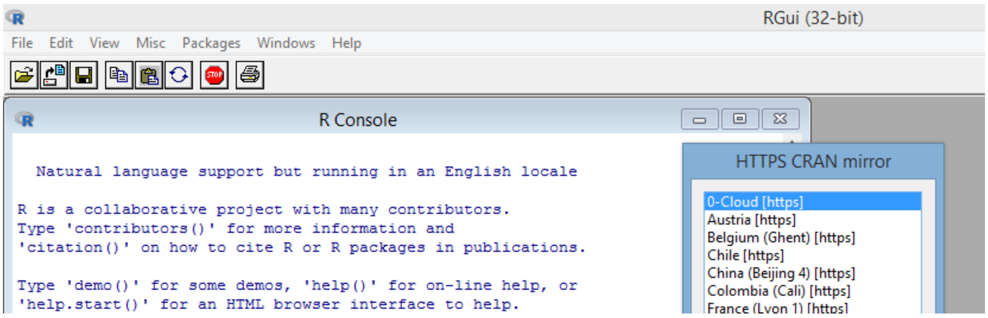
4.2.1 Step Analysis with *R* program

Support vector regression testing is used to predict productivity loss in pile construction productivity in Indonesia. It is carried out by several phases, namely the sharing of data, parameter optimization, and testing. Data sharing is done using the input of attributes and instances into ***R*** software versions below 3.3.2 , equipped with interfaces and extensions by SVM. Used dataset were consisting of 5 attributes and 110 project. These are the step that used to appearance of ***R*** software:

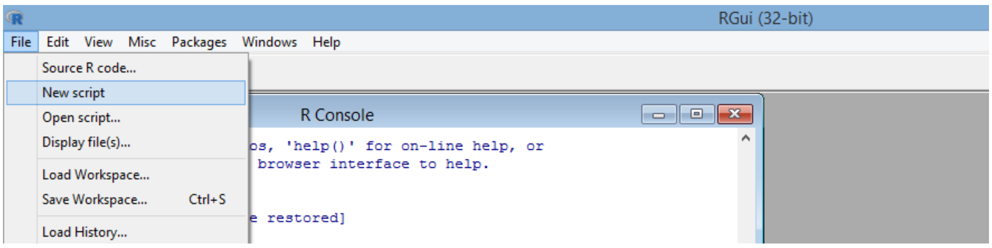
a) Download ***R*** versions below 3.3.2

b) Open ***R***, select the package → install package (s) → 0-cloud (https) → e1071 (this is the package for the analysis of SVM / SVR)





c) Select the file → select new scrip →copy syntax to new script on a new script



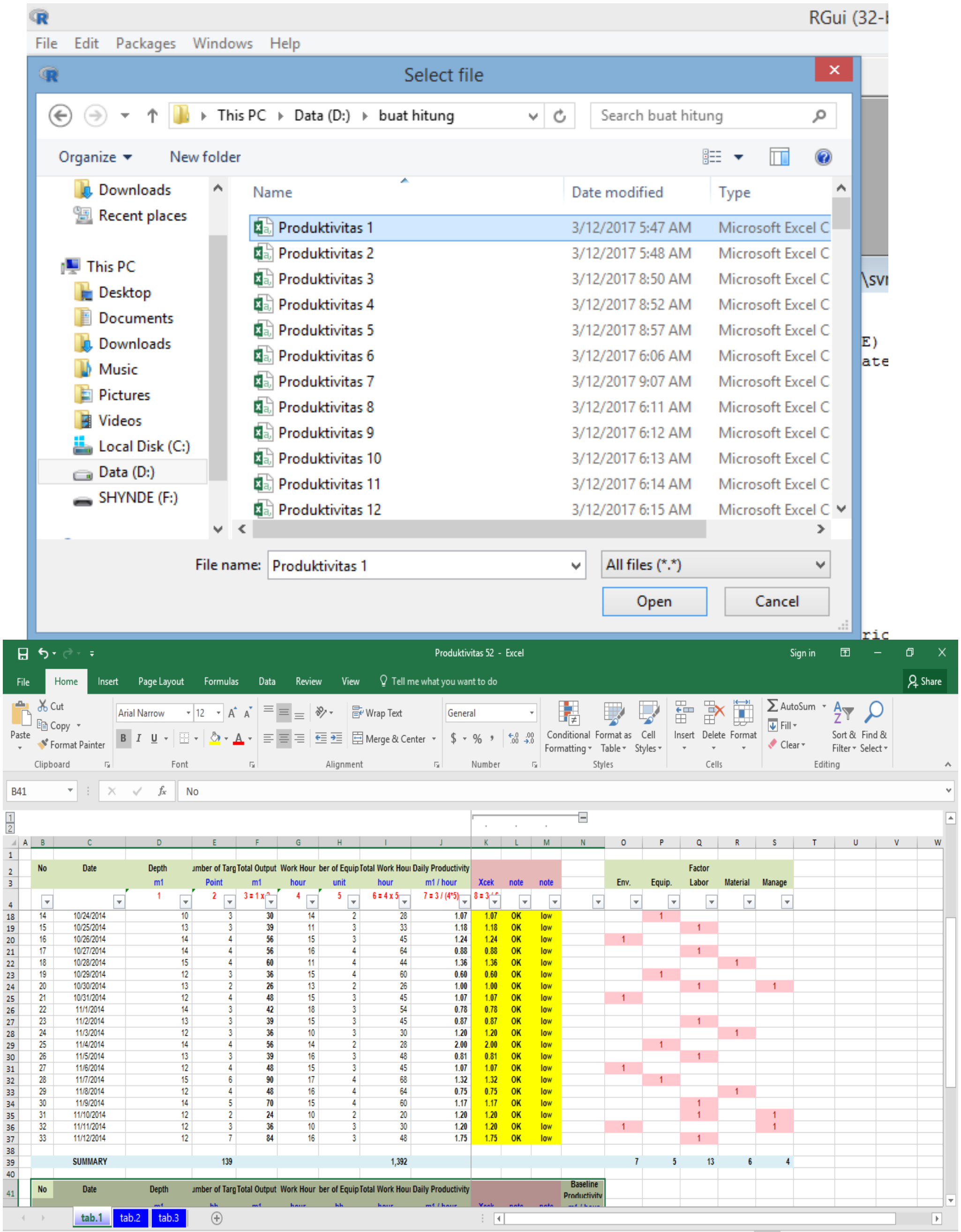
d) Syntax Explaination

```
# to call the package
library(e1071)

#to call the data
data1 <- read.csv(file.choose(), header=TRUE)

#model svr
model1<-
svm(loss~pileDepth+Numberofpile+sizeproject+numberofequipment+location+typeofpilr+soilcondit
ion+city, data1)

#make prediction/ forecasts, model1 is estimating model parameters svr dan data1 is data used
for prediction
predict1<-predict(model1,data1)

#separating the actual value of the loss productivity to calculate the error and create graphics
actual1<-data1[,6]

#look for the number of data row
n<-nrow(data1)

#to find the number of periods of the data
x<-seq(1,n,1)

#calculating error
e<-actual1-predict1

#calculating rmse
rmse<-(sum(e^2)/n)^(1/2)

#calculate the correlation
correlation<-cor(aktual1,predict1)

#displays of model parameters, RMSE and correlation
model1
rmse
correletion

#making plot
#create a plot for the actual data
```

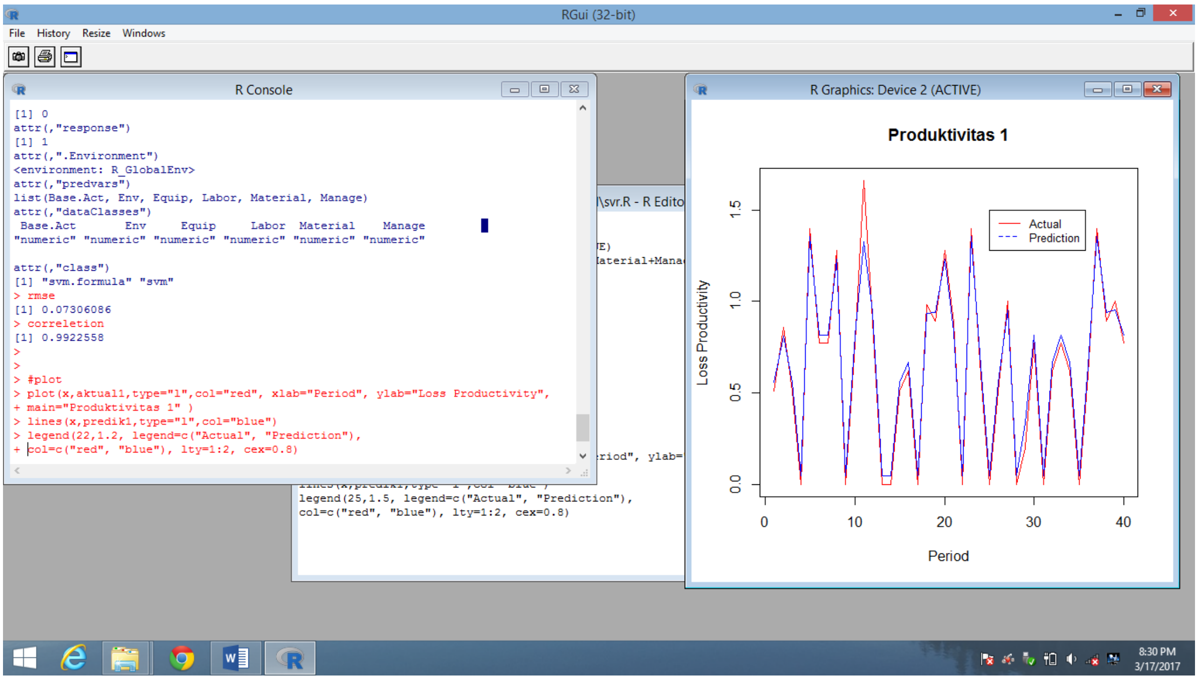
```
plot(x,actual1,type="l",col="red", xlab="Period", ylab="Loss Productivity",main="Produktivitas 20" )
#add data prediction
lines(x,predik1,type="l",col="blue")
#add a legend
legend(22,1.2, legend=c("Actual", "Prediction"),col=c("red", "blue"), lty=1:2, cex=0.8)
```

e) How to perform syntax by pressing ctrl+r on reach line

- ✓ Run the first package
- ✓ After the completion of the run press ctrl + r call up data, appear as below select data used → click OK
- ✓ The data used for the prediction that table on **tab 1 in the excel program**



✓ Block models SVR until the end, press crtl+r



✓ Result

This model create the predictive model by analyzing linear shape patterns between the factors and construction productivity. The numeric results indicate projects that have individual attribute data included all influencing factors and construction productivity. When the projects have two different patterns regarding attribute information, these patterns cannot be stated by a general regression analysis method. As a result, Support Vector Machines are employed in this case and Support Vector Regression Model discover patterns for percentage accuracies. While the non-linear predictive pattern of percentage accuracies were unable to be solved in two dimensional spaces, the non-linear plotting is used to plot the data into a high dimensional feature space in which linear regression is presented.

The graphics presenting predictive pattern of percentage accuracies of project cost estimation is transformed into a linear shape. Lastly the predictive model, accomplished to predicting the percentage accuracies even if data is filed in two dimensional spaces, is established. Also, Support Vector Machines can be employed to regression illustrations by the beginning of a loss operation. They apply classifiers of an adaptable flexibility, which can lead a better accuracy. It is because overfitting is prevented by adapting error range and classifiers are optimized on the training data for a high-quality generalization achievement.

In Support Vector Regression, the graphics performing predictive pattern of percentage accuracies is explained by equation. The factor coefficient is a fixed number that impacts the independent factor, which resulting the percentage accuracies.

4.3 Interpretation Prediction Result and Discussion

SVR Model is a model of SVM that can be used to predict the productivity within the next period. To view the performance of the model in this study. Root Mean Square Error (RMSE) was used. The smaller number of RMSE performances, the better the number of the produced model. From 110 data pile construction productivity in this study, the best performance or the smallest RMSE. There are many productivity that has big RMSE value (more than 1), because there is an adequately high value that are different from the others because it has a big margin. Although the RMSE value is high, but the value of correlation remains high, because the predicted results still consistent with the pattern of actual data. Data prediction has the same pattern with the actual data, so that the model is good enough to be used for the prediction. Overall (productivity 1- 110) have the same model with the SVR parameters

c	=	*1*
gamma	=	*0.125*
epsilon	=	*0.1*

Therefore, it can be concluded that the model is good enough to be used as a prediction tools.

In SVM calculation, a penalty was given by adding cost C value. Cost C value were chosen in order to control the balance between the margin and error prediction. The bigger C value is, the bigger the given penalty.

Data mining assumption is linearity. Therefore, the produced algorithm was limited for linear cases. If a case shows irregular data, algorithm like perceptron can’t overcome it. Generally, real world cases are un-linear case. If such case happened, kernel method is a method to be used. By using kernel method, data x in input space may be mapped into feature space F with higher dimension.

The kernel function that’s going to be used in substituting dot product in feature space is highly dependent on the data. Usually, cross-validation method was used in choosing this kernel function. Deciding on kernel function is a very important step, because this kernel function also decides which feature space is going to be used in order to determine the prediction function (Nugroho, 2008).

Some of the actual data in this study (loss productivity) have zero value which caused the value to be incalculable, so the level degree of the best model result can we see in the RMSE and correlation value. This percentage of error rate can not be calcuates , because of the formula are :

$$Error\ Rate \quad = \quad \frac{actual\ data - prediction\ result}{actual\ data}\ X\ 100\% \qquad \textbf{(7)}$$

The best performance or the smallest Root Mean Square Error is 0.01 with squared correlation coefficient rate 0.87.

```
> model1

Call:
Svm
(formula=loss~pileDepth+Numberofpile+sizeproject+numberofequipment+location+typeofpilr+soilcondition
+city, data = data1)



Parameters:
   SVM-Type:  eps-regression
 SVM-Kernel:  radial
       cost:  1
      gamma:  0.125
    epsilon:  0.1



Number of Support Vectors:  22

> rmse
[1] 0.01078526
> correlation
[1] 0.8722584
```

Out of 110 projects, 100 projects data has been used for training the network (90 % of total data) and remaining10 projects data (10%) has been used for testing. During training the network has predicted the production rates with lower values of MSE and follows similar trend and pattern of target values as shown in **Figure 4.2.** The prediction graph in result with higher RMSE value still follow the pattern of the actual value graph because it still has high correlation value. Y-axis represent the value of productivity loss from a factor during construction, which was taken during 40 days of work and X-axis represents the time.

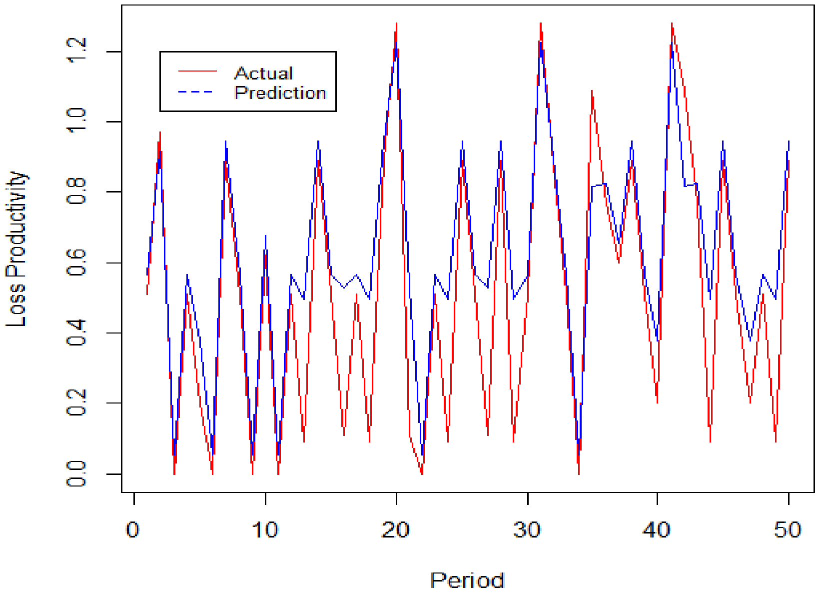


Figure 4.2 Curve Output Training Error

Productivity rates values predicted during testing also have lower error values and follows almost similar trend and pattern with slight variation at the end as shown in **Figure 4.3**.

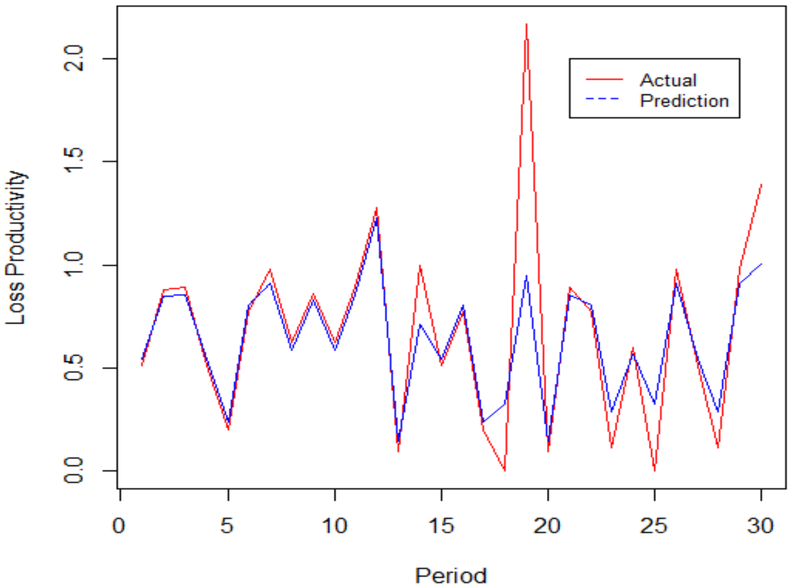


Figure 4.3 Prediction Testing Curve Result

Average values of each project predicted rates have been calculated then the MSE of training and testing predicted rates have been determined for training and testing as shown in **Table 4.7** and **Table 4.8**. Average errors calculated for training is 0.87 whereas testing error are slightly lower than training error with values 0.011. Then average of MSE has been calculated of all the projects. The results shows that the error of training and testing of the actual production rates and also the values calculated for MSE of training and testing is shows that the prediction has achieved better convergence.

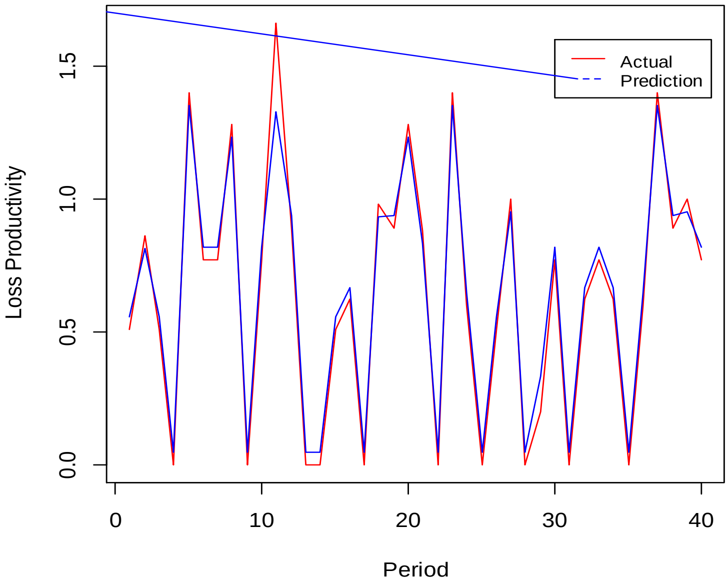


Figure 4.4 Prediction Graph Result

From the **Figure 4.4**, the predicted results curved patern still consistent with the pattern of actual data. Data prediction has the same pattern with the actual data, so that the model is good enough to be used for the prediction, the paterns follows similar trend and pattern of target values. The prediction graph in result with low RMSE value still follow the pattern of the actual value graph because it still has high correlation

Table 4.7 Model Result Traning Data

Combination 10-fold cross-validation with Data Training	Total Mean Squared Error	Squared Correlation Coefficient (Accuracy Rate)
1	0.11	0.68
2	2.06	0.07
3	2.15	0.42
4	2.1	0.41
5	0.06	0.75
6	0.00	0.99
7	0.07	0.73
8	0.01	0.72
9	0.01	0.73

Combination 10-fold cross-validation with Data Training	Total Mean Squared Error	Squared Correlation Coefficient (Accuracy Rate)
10	2.17	0.8
AVERAGE	0.87	0.63

Table 4.8 Model Result Testing Data

Combination 10-fold cross-validation with Data Testing	Total Mean Squared Error	Squared Correlation Coefficient (Accuracy Rate)
1	0.01	0.74
2	0.02	0.83
3	0.01	0.87
4	0.01	0.94
5	0.01	0.91
6	0.01	0.87
7	0.02	0.82
8	0.01	0.92
9	0.01	0.89
10	0.01	0.90
AVERAGE	0.011	0.868

Table 4.9 Difference Result

	Total Mean Squared Error	Squared Correlation Coefficient (Accuracy Rate)	Note
Combination 10-fold cross-validation with Data Training	0.87	0.63	The model is adequate if it had high correlation value and smaller MSE value
Combination 10-fold cross-validation with Data Testing	0.011	0.868	
Difference	0.859	-0.238	

Accuracy of the prediction can be seen from the **Figure 4.3** and **Figure 4.4**, the pattern and correlation between the actual value and the prediction. The model is adequate if it had high correlation value and smaller MSE value. For further research, existing parameter may be inserted into the model as well as the impacting factor, which will result the prediction. There are three parameters that were used for SVR, those are cost, gamma, and epsilon. Success rate may be seen from the error rate, however, because the actual y-axis contains 0 value, the error rate can’t be calculated.

This model create the predictive model by analyzing non - linear shape. Optimal productivity is the highest achievable sustainable productivity under good management and typical field conditions. Optimal productivity is useful in determining the absolute construction operations efficiency because an accurate estimate of optimal pile construction productivity allows the comparison of actual vs. optimal productivity.

Research variable was composed from response variable (y) and predictor variable (x). Initial data processing was done using Microsoft Excel by grouping x and y variable. Data processing using Support Vector Machine was used for measuring pile productivity loss in the future / target (y), and the used input data were 110 data set.

For parameter setting using SVM algorithm, it is known that productivity loss configuration prediction was by using data input of the previous project for k-fold 10, C (cost) = 1 and kernel type radial. Configuration design to predict the productivity loss for the future was calculated, and the result is:

Table 4.10 Prediction Result Recapitulation with 8 Factors

No.	Total Mean Squared Error	Squared Correlation Coefficient (Accuracy Rate)	Number of Support Vectors	Parameter	K = 10 fold				
1	0.01	0.74	16.00	Cos = 1 Gamma = 0.125 Epsilon = 0.1	0.105	0.002	0.000	0.001	0.003
					0.001	0.002	0.000	0.000	0.000
2	0.02	0.83	23.00		0.130	0.001	0.006	0.022	0.001
					0.001	0.006	0.007	0.012	0.001

No.	Total Mean Squared Error	Squared Correlation Coefficient (Accuracy Rate)	Number of Support Vectors	Parameter	K = 10 fold				
3					**0.022**	**0.001**	**0.001**	**0.037**	**0.010**
	0.01	**0.87**	**22.00**		**0.006**	**0.004**	**0.005**	**0.002**	**0.004**
4					0.001	0.005	0.012	0.001	0.005
	0.01	0.94	21.000		0.019	0.004	0.001	0.000	0.005
5					0.001	0.001	0.001	0.009	0.014
	0.01	0.91	20.00		0.037	0.004	0.004	0.007	0.001
6					0.005	0.022	0.008	0.000	0.001
	0.01	0.87	18.00		0.010	0.004	0.038	0.002	0.010
7					0.001	0.006	0.140	0.014	0.010
	0.02	0.82	23.00		0.001	0.001	0.006	0.002	0.001
8					0.005	0.000	0.006	0.017	0.001
	0.01	0.89	20.00		0.013	0.019	0.001	0.008	0.002
9					0.006	0.004	0.004	0.013	0.006
	0.01	0.92	23.00		0.007	0.005	0.008	0.007	0.006
10					0.000	0.001	0.020	0.001	0.004
	0.01	0.90	20.00		0.044	0.016	0.001	0.003	0.002

From the **Table 4.10** the best performance or the smallest Mean Square Error is **0.01** with Squared Correlation Coefficient rate **0.87** with **22.00** number of support vector. Data processing in this research was by using SVM algorithm with data calculation construction: data calculation using R program, which was processed by inputting data (10 data combination) as training data, followed by deciding kernel type, C (cost) value as well as the k-fold. The following are the explanation for determining SVR parameter:

Cos	=	1 cost of constraint violation (default: 1). This is the 'C'-constant of the regularization term in the Lagrange formulation
Gamma	=	0.125 parameter needed for all types of kernels except linear, Default : 1/(data dimension) , Data dimension : 8 factor
Epsilon	=	0.1 (Default)

Type Kernel defines the kernel used in training and prediction. The options are: linear, polynomial, radial basis and sigmoid. The kernel type that were used was radial and the C (cost) value = 1, and k-fold number was k = 10. The following are the testing result that were done using several kernel functions and inserting C (cost) value as well as range (k-fold) value that were determined on each data set. After determining k-fold validation, c(cost) and kernel type, smallest MSE (mean square error) were determined. The said smallest MSE is the one that were used as design to predict as well as accuracy benchmark.

In **Table 4.11**, SVR models also provides predictions about the potential loss of input faced by the project, from the model, it can be seen the potential number of prediction loss. Looping being performed for 40 times to get the most similar number. The potential value for measuring good productivity loss is the smallest and the most repeatable value of the prediction itself and the influence of the 8 factors used in this research related to the effect of pile construction productivity performance itself. One of the method for calculating the similarity is Euclidean distance.

The formula is as follows:

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}. \qquad (8)$$

Table 4.11 Sample of Prediction Potential Loss Result

Comparison Data Number	Work Hour Lost in Data Model	Work Hour Lost from Comparison Data	Euclidean Distance
16	**76.61**	**76.65**	**0.04**
53	76.61	76.89	0.28
24	76.61	76.26	0.35
97	76.61	76.18	0.43
60	76.61	76.18	0.43
66	76.61	76.06	0.55
99	76.61	77.76	1.15
4	76.61	77.80	1.20
35	76.61	75.33	1.28
9	76.61	75.24	1.37

Based on **Table 4.11**, the most similar case is case number 16 because it has the smallest number of euclidian distance with the model.

Table 4.12 Recapitulation of Potential Loss Result

Lost Productivity (hour)	Environment	12.91
	Equipment	66.68
	Labor	58.21
	Material	16.18
	Management	76.65
Total Work Hour Lost		**230.65**
Overall work hour		1243
Productivity Loss Percentage		**0.1855 (18.55%)**

From the results in the **Table 4.12,** it can be seen the potential loss prediction is **0.1855 (equivalent to 18.55%)**, so it can be concluded that this model can be used to predict the productivity loss of a project in the future, and forecasting using SVR method has the smallest standard error value and the results is close to the original data, potential loss prediction aims to reduce lost work hours caused by factors occur in the project. In the future it aims to improve productivity.

Productivity is commonly defined as the quantity of work produced of work output per unit of input or effort (Klanac, 2004). Loss in productivity is usually experienced by contractor and entails the completion of work at less than the planned rate of production. There has been a considerable amount of research about the factor that causes loss in productivity common factors include acceleration (directed or constructive) adverse or unusually severe weather, the cumulative impact of multiple changes and revisions, site or work area access restrictions, site conditions, untimely approvals and responses to labour market conditions (Klanac, 2004).

From the result done of major activities the outcome related to loss in productivity and overall factors incurred gives a new dimension for engineer to study the delay causes with loss in productivity. Correlation of loss in productivity with factors causes should be given more importance in the building construction trade by good management. Therefore real factors causes can be summarized as a loss in productivity problem to reflect the impact on productivity loss.

CHAPTER V: CONCLUSION

5.1 Conclusion

The uniqueness of construction projects and the inherent complexity of this industry make it indispensable to set the dynamic performance indexes applied to each task. Most of the methods for measuring productivity and determining the standard basis are holistic. Such techniques do not consider net sources in determining the policy and benchmarks, and consequently are selecting a part of project performance as the foundation. The objective-based access and definition of productivity measurement methods, introduction of net baseline productivity index and its comparison with macro-level productivity of the project, Its main advantage is the exclusion of resource losses in determining base value compared to other available methods. Furthermore, the real features and benefits are proven by full implementation of the framework in the shape of the proposed methodology in the analyzed cases.

The new approach to baseline productivity is one of the other outcomes of this research; this approach is described in details by focusing on macro impact factors pile construction productivity. This approach is applicable to the machinery, tools and materials, and its potential and scientific applications can be indicated by their experimental implementation on case studies.

The various values of net baseline productivity in different elements of a project and comparing them with reference values provided by formal systems can be elicited as the other subjects for future studies.

The investigation of performance and impact of this methodology in regard to the corresponding parties involved in the projects such as the client or supervising body is the issues that can be seen in future works.

The calculations for the loss of value of lost productivity and work hours from 110 project (Total Workday = 175.601 hours) that there are for lost working hours, with details as follows:

a. Factors Equipment raises work hours lost by 7828,07 hours
b. The factors causing Labor work hours lost by 2422,75 hours
c. Factors causing work hours lost Management by 1574,14 hours

Macro level factor of pile construction productivity is Equipment the biggest factor that affect project's efficiency and productivity. However, because the said factors usually are not specific, it was usually can be calculated from economic point of view only. By using SVR (Support Vector Regression) , we may predict productivity loss by determining which factor is required to be reduced or anticipated. All of which is related to the regulation, as well as site location and factor correlation.

Factors influencing these rates such as weather, availability of material and equipment, location of project, site conditions and number of workers which are subjective in nature, have been recorded on scale at sites. To determine the individual effect and severity level of each factor severity indices have been calculated. Availability of the materials and equipment is the most severe factor identified that indicates that improper management of materials and equipment handling has greater influence on the accurate estimation of productivity rates.

Reliable values of production rates with incorporation of these factors have been successfully predicted by SVR model. Performance of the model has been determined by calculating the percentage error and MSE (Mean Square Error) of the predicted production rates. From 110 data pile construction productivity in this study, the model was evaluated and implemented using the 10-folds cross validation to develop SVR model in order to predict productivity loss in pile construction productivity in Indonesia.. The error values of **0.87** and **0.63** are the values calculated for MSE of training and testing outputs whereas correlation **0.011** and **0.868** have been obtained for training and testing output results. The best performance or the smallest Mean Square Error is **0.01** with Squared Correlation Coefficient rate **0.87** with **22.00** number of support vector and accuracy rate **87,25931%** and potential productivity loss **18.55%.** These results indicate that the SVR model has predicted production rates values for concreting in columns reasonably within acceptable range of errors.

5.2 Research Recommendation

The research has produced optimum and accurate prediction, however, in order to produce better result on the next research, these following subjects are required to be improved:

1. More amount of observation data for simulation is required, in order to produce more data training process model thus resulting in prediction with higher accuracy. Furey (2000) stated that with higher amount of data, SVM may be tested on experiment with higher scale, therefore the validation from the comparison result can’t be doubted.
2. In order to produce better prediction, an optimization for SVM is required by using linear function within a feature space and implementing learning bias.

REFERENCE

AECOM. 16/04/2014). Asia Construction Outlook 2014 Released. *Asia Outlook*.

Atkinson, W., Wells. (1997). A Stakeholderpproach to Strategic Performance Measurement. *Sloan Management Review, 38*(3), 237.

Bank, W. (2009). Towards 2015 : Spending for Indonesia's Development, Shaping the Prospects of a Middle-Income Country.

Box, G. E. P., dan G. M. Jenkins. (1976). *Time series Analysis, Forecasting, and Control*. San Fransisco: Holden-Day.

Durdyev, S., & Mbachu, J. . (2011). On-site Labour Productivity of New Zealand Construction Industry: Key Constraints and Improvement Measures. *Australasian Journal of Construction Economic and Building, 3*, 18 - 33.

Enshassi., A., Sherif Mohamed, Saleh Abushaban. (2009). FACTORS AFFECTING THE PERFORMANCE OF CONSTRUCTION PROJECTS IN THE GAZA STRIP. *JOURNAL OF CIVIL ENGINEERING AND MANAGEMENT*, 269-280. doi:10.3846/1392-3730.2009.15.269-280

Gidado, N. A. a. K. (2009). Evaluation of Factors Affecting Productivity in the UAE Construction Industry: Regression Models. 12.

Groak, S. (1994). Is construction an industry? Notes towards a greater analytic emphasis on external linkages. *Construction management and economics, 12*, 287-293.

Halpin, T. M. Z. a. D. W. (2004). Quantitative Assessment for Piles Productivity Factors. *Journal of Construction Engineering and Management*, 13. doi:10.1061/(ASCE)0733-9364(2004)130:3(405)

Han, J. (2006). *Data Mining : Concepts and techniques*: Morgan Kaufman.

Hanna, A. S., Russell, J. S., Gotzion, T. W., and Nordheim, E. V. (1999a). Impact of change orders on labor efficiency for electrical construction. *Journal of Construction Engineering and Management, Vol. 125-3*, 8.

Homyun Jang, K. K., Juhyung Kim, and Jaejun Kim (2011). Labour productivity model for reinforced concrete construction projects. *Construction Innovation Process,, 1*, 11.

Huang. (2009). Metrics and Tools for Measuring Construction Productivity: Technical and Empirical Considerations. *NIST Special Publication, 1101*.

Industry, C. o. C. (2014). CBSL Annual Report. *Institute of Construction Training and Development*.

Jan Bröchner, A. A. S. (2013). Construction productivity growth and business services. *Economies et societes, 47*(3-4), 509-529.

Joachims, T. (2002). *Learning to classify text using support vector machines: Methods, theory and algorithms*: Kluwer Academic Publishers.

Jonsson, J. (1996). *Construction site productivity measurements: selection, application and evaluation of methods and measures.* (Ph.D), Luleå University of Technology, Sweden.

Klanac, G. P. a. N., E. L. (2004). Trends in construction Loss in productivity claims. *Journal of Professional Issues in Engineering Education and Practice, Vol. 130*(No. 3).

Koehn, E. a. B., G. , & (1985). Climate effects on construction. *Journal of Construction Engineering and Management, Vol. 111-2*, 8.

Lam, K. C., Palaneeswaran, E., & Yu, C.-y. (2009). A support vector machine model for contractor prequalification. *Automation in Construction, 18*(3), 321-329.

Lesmana., A. P., Ratna S. Alifen. (2014). ANALISIS PRODUKTIVITAS PEKERJAAN PONDASI BORED PILE (STUDI KASUS PADA BANGUNAN PERKANTORAN 31 LANTAI). 7.

Makridakis, S., Wheelwright, SC, and Hyndman, RJ (1998). *Forecasting: methods and applications.* (Vol. 3rd edition). New York.

Mawdesley M J, Q. S. (2000). *Systems Thinking and Construction Productivity, .* Paper presented at the International Conference on Systems Thinking in Management: Dynamics of Theory and Practice, Geelong, Victoria, Australia.

Naguib, I. A. a. H. W. D. (2012). *Support vector regression and artificial neural network models for stability indicating analysis of mebeverine hydrochloride and sulpiride mixtures in pharmaceutical preparation: A comparative study. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*.

Nugroho. (2008). *SVM: Paradigma Baru dalam Softcomputing dan Aplikasinya*. Paper presented at the Konferensi Nasional Sistem & Informatika 2008, Bali.

Oglesby, C. H., Parker, H. W., and Howell, G. A. (Ed.) (1989). *Productivity Improvement in Construction*.

Olomolaiye P, J. A., Harris F (1998). Construction Productivity Management. *Chartered Institute of Building, UK.*

Park, H.-S. (2006). Conceptual framework of construction productivity estimation. *KSCE Journal of Civil Engineering, 10*(5).

Park, H.-S., Thomas, S.R. & Tucker, R.L. (2005). Benchmarking of Construction Productivity. *Journal of Construction Engineering and Management, 131*(7), 772-778.

Parrella, F. (2007). Online support vector regression. *Master's Thesis, Department of Information Science, University of Genoa, Italy*.

Paul W. Chan, A. K. (2007). Productivity improvements: understand the workforce perceptions of productivity first. *Personnel Review, 36*(4), 564 – 584.

Pekuri, A., Haapasalo, H., & Herrala, M. (2011). *International Journal of Performance Measurement, 1*, 19-58.

Peurifoy, R. L., Ledbetter, W. L., and Schexnayder, C. J. . (1996). *Construction planning, equipment, and methods* (5 ed.). New York: McGraw-Hill.

PF Kaming, G. H., ST Kometa, and P Olomolaiye. (1998). Severity diagnosis of productivity problems – a reliability analysis. *International Journal of Project Management, 16*(2), 107–113.

Sanders, S. R. a. T., H. R. (1993). Masonry productivity forecasting model. *Vol. 119-1*, 16.

Santosa, B. (2010). Tutorial Support Vector Machine. *Teknik Idustri, ITS.[Online]. Tersedia:* http://www. *google. co. id/url*.

Scholkopf, B. a. A. J. S. (2001). *Learning with kernels: support vector machines, regularization, optimization, and beyond*: MIT press.

Shah, R. S. (2007). *Support Vector Machines for Classification and Regression.* McGill University.

Smith, S. D. (1999). Earthmoving productivity estimating using linear regression techniques. *Vol. 125-3*, 8.

Soekiman., A., Krishna S. PRIBADI,Biemo W. SOEMARDI,Reini D. WIRAHADIKUSUMAH. (2011). STUDY ON FACTORS AFFECTING PROJECT LEVEL PRODUCTIVITY IN INDONESIA. *ANNALS OF FACULTY ENGINEERING HUNEDOARA - INTERNATIONAL JOURNAL OF ENGINEERING, 3*, 6.

Stainer, A. (1997). Capital input and total productivity management. *Management Decision*, *35*(3), 224 - 232.

Subagyo, P. (1986). *Forecasting Konsep dan Aplikasi*. Yogyakarta: BPFE UGM.

Sumanth, D. J. *Total Productivity Management (TPmgt): A Systemic and Quantitative Approach to Compete in Quality, Price and Time*: CRC Press.

Sweis, G. J. (2000). Impact of conversion technology on productivity in masonry construction.

Thomas, H. R., Mathews, C. T., and Ward, J. G. (1986). “Learning curve models of construction productivity. *Journal of Construction Engineering and Management, Vol. 112-2*, 13.

Thomas, H. R. a. Y., I. (1987). Factor model of construction productivity. *Journal of Construction Engineering and Management, Vol. 113-4*, 15.

Thomas, H. R. a. Z., I. (1999). Construction baseline productivity: Theory and practice. *Journal of Construction Engineering and Management, Vol. 125-5*, 8.

Varma Santosh, M. R. A. (2014). Productivity in Building Construction. *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE), 10*(5), 64-71.

Wang, Y.-R., C.-Y. Yu, and H.-H. Chan. (2012). Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models. *International Journal of Project Management, 4*, 30.

Winch, G. a. C., B (2001). Benchmarking on-site productivity in France and the UK: a CALIBRE approach. *Construction management and economics, 19*, 577-590.

Yasin, H., Prahutama, A., & Utami, T. W. (2014). PREDIKSI HARGA SAHAM MENGGUNAKAN SUPPORT VECTOR REGRESSION DENGAN ALGORITMA GRID SEARCH. *MEDIA STATISTIKA, 7*(1), 29-35.

APPENDIX : PARAMETER RESULT

a. Summary Productivity Loss caused by Macro Impact Factor

No	Work Start Date	Pile Depth (m)	Work Finish Date	Total Number of Pile Finished (quantity)	Total Work Hours	Work Hour Loss from Macro Impact Factor	Macro Impact Factor (Productivity Loss Caused by)
1	9/15/2014	37	10/21/2014	147	1,653	76.61	Equip.
2	10/10/2011	30	11/8/2011	125	1,177	95.78	Equip.
3	5/4/2011	30	6/2/2011	146	1,165	64.69	Equip.
4	6/11/2011	30	7/10/2011	150	1,331	77.80	Manage
5	8/7/2011	35	9/10/2011	165	1,590	88.42	Equip.
6	7/8/2011	37	8/13/2011	180	1,588	90.16	Labor
7	8/10/2011	37	9/15/2011	174	1,677	54.87	Labor
8	8/13/2011	37	9/18/2011	180	1,650	55.75	Labor
9	10/22/2011	40	11/30/2011	184	1,782	75.24	Labor
10	11/2/2011	50	12/21/2011	224	2,252	130.22	Labor
11	12/5/2011	50	1/23/2012	234	2,185	322.40	Equip.
12	12/16/2011	40	1/24/2012	236	1,719	122.22	Manage

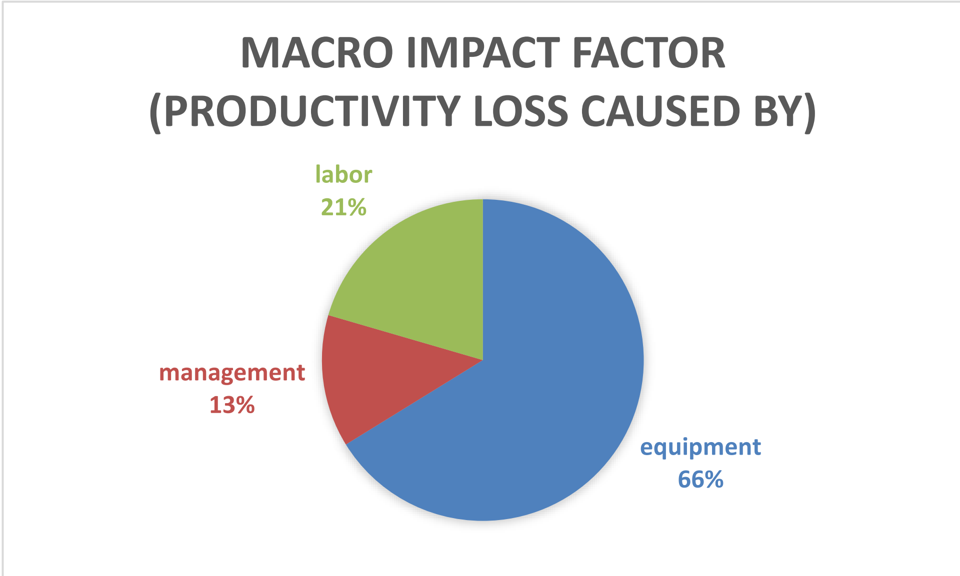
13	12/21/2011	40	1/29/2012	223	1,626	182.29	Equip.
14	1/5/2012	35	2/8/2012	206	1,471	124.69	Equip.
15	3/14/2012	35	4/17/2012	268	1,574	87.06	Equip.
16	4/16/2012	28	5/13/2012	213	1,243	76.65	Manage
17	5/3/2012	24	5/26/2012	187	1,040	70.13	Manage
18	6/2/2012	30	7/1/2012	231	1,322	92.55	Equip.
19	7/13/2012	40	8/21/2012	181	1,760	119.50	Equip.
20	8/17/2012	30	9/15/2012	173	1,187	61.61	Labor
21	8/11/2012	28	9/7/2012	204	1,243	99.28	Manage
22	8/25/2012	50	10/13/2012	260	1,719	203.25	Equip.
23	10/11/2012	35	11/14/2012	168	1,590	153.90	Equip.
24	10/29/2012	30	11/27/2012	152	1,331	76.26	Equip.
25	12/22/2012	35	1/25/2013	180	1,556	99.77	Labor
26	1/2/2013	35	2/5/2013	210	1,471	107.47	Equip.
27	2/17/2013	37	3/25/2013	175	1,677	71.37	Manage
28	4/11/2013	40	5/20/2013	229	1,659	175.51	Equip.
29	6/17/2013	50	8/5/2013	233	2,252	98.11	Equip.
30	7/10/2013	40	8/18/2013	235	1,689	106.10	Equip.
31	9/6/2013	34	10/9/2013	165	1,565	97.61	Labor
32	11/3/2013	30	12/2/2013	144	1,159	92.78	Equip.
33	12/13/2013	30	1/11/2014	150	1,381	91.20	Manage

34	1/6/2014	28	2/2/2014	200	1,260	96.15	Equip.
35	1/22/2014	30	2/20/2014	250	1,181	75.33	Equip.
36	2/13/2014	50	4/3/2014	232	2,185	139.01	Equip.
37	2/28/2014	50	4/18/2014	203	2,167	174.76	Labor
38	3/9/2014	40	4/17/2014	230	1,711	116.29	Manage
39	4/11/2014	24	5/4/2014	183	1,040	60.99	Labor
40	4/20/2014	28	5/17/2014	216	1,243	94.56	Equip.
41	5/2/2014	34	6/4/2014	162	1,560	102.82	Labor
42	5/17/2014	33	6/18/2014	156	1,485	99.51	Manage
43	6/2/2014	39	7/10/2014	231	1,674	172.80	Equip.
44	6/15/2014	30	7/14/2014	223	1,277	99.05	Equip.
45	7/13/2014	35	8/16/2014	168	1,584	154.08	Equip.
46	7/28/2014	28	8/24/2014	208	1,243	56.80	Manage
47	8/3/2014	35	9/6/2014	180	1,590	128.43	Labor
48	8/27/2014	50	10/15/2014	222	2,252	112.61	Equip.
49	9/5/2014	40	10/14/2014	230	1,677	130.43	Equip.
50	9/22/2014	30	10/21/2014	146	1,331	93.87	Manage
51	10/3/2014	42	11/13/2014	198	1,832	155.88	Labor
52	10/11/2014	33	11/12/2014	139	1,392	150.88	Labor
53	10/25/2014	37	11/30/2014	177	1,677	76.89	Manage
54	11/6/2014	50	12/25/2014	221	2,198	119.25	Equip.

55	11/18/2014	30	12/17/2014	146	1,177	92.78	Equip.
56	12/4/2014	30	1/2/2015	140	1,301	90.93	Manage
57	12/30/2014	50	2/17/2015	237	2,185	171.91	Equip.
58	1/8/2015	35	2/11/2015	214	1,471	72.34	Equip.
59	1/16/2015	40	2/24/2015	229	1,662	96.13	Equip.
60	1/30/2015	30	2/28/2015	252	1,185	76.18	Equip.
61	2/4/2015	35	3/10/2015	171	1,590	117.43	Equip.
62	2/20/2015	40	3/31/2015	235	1,728	97.89	Equip.
63	3/16/2015	37	4/21/2015	172	1,677	55.72	Labor
64	3/22/2015	35	4/25/2015	215	1,471	126.36	Equip.
65	4/4/2015	35	5/8/2015	272	1,574	67.87	Equip.
66	4/15/2015	40	5/24/2015	185	1,803	76.06	Labor
67	5/12/2015	30	6/10/2015	149	1,179	60.32	Equip.
68	5/16/2015	50	7/4/2015	225	2,252	123.74	Equip.
69	6/8/2015	40	7/17/2015	262	1,719	97.15	Equip.
70	6/14/2015	50	8/2/2015	227	2,121	147.14	Equip.
71	7/2/2015	40	8/10/2015	226	1,651	128.17	Equip.
72	7/19/2015	30	8/17/2015	153	1,144	78.13	Equip.
73	8/2/2015	40	9/10/2015	241	1,730	117.64	Equip.
74	8/13/2015	34	9/15/2015	167	1,565	81.73	Labor
75	8/23/2015	30	9/21/2015	155	1,114	49.21	Equip.

76	9/6/2015	30	10/5/2015	148	1,331	99.75	Manage
77	9/17/2015	28	10/14/2015	209	1,243	79.33	Equip.
78	10/13/2015	30	11/11/2015	254	1,193	45.66	Equip.
79	10/21/2015	50	12/9/2015	238	2,185	221.38	Equip.
80	10/28/2015	37	12/3/2015	176	1,677	60.83	Labor
81	11/11/2015	32	12/12/2015	186	1,329	67.94	Labor
82	11/20/2015	50	1/8/2016	235	1,719	140.66	Equip.
83	12/5/2015	34	1/7/2016	164	1,560	139.40	Labor
84	12/13/2015	33	1/14/2016	156	1,527	137.49	Labor
85	1/4/2015	35	2/7/2015	274	1,574	80.69	Equip.
86	1/23/2015	40	3/3/2015	235	1,853	105.82	Labor
87	2/8/2015	50	3/29/2015	182	2,252	209.90	Equip.
88	2/13/2015	35	3/19/2015	167	1,590	106.65	Equip.
89	3/5/2015	40	4/13/2015	264	1,719	83.20	Equip.
90	3/19/2015	40	4/27/2015	266	1,719	85.12	Manage
91	4/7/2015	40	5/16/2015	239	1,719	102.92	Equip.
92	4/22/2015	35	5/26/2015	186	1,471	111.29	Equip.
93	5/6/2015	35	6/9/2015	169	1,590	139.48	Equip.
94	5/17/2015	28	6/13/2015	138	951	79.05	Manage
95	6/2/2015	30	7/1/2015	150	1,177	73.09	Equip.
96	6/13/2015	40	7/22/2015	242	1,719	109.01	Equip.

97	7/9/2015	40	8/17/2015	188	1,803	76.18	Manage
98	7/15/2015	40	8/23/2015	244	1,719	165.00	Equip.
99	8/3/2015	40	9/11/2015	246	1,719	77.76	Manage
100	8/22/2015	40	9/30/2015	170	1,590	172.47	Equip.
101	9/12/2015	50	10/31/2015	228	2,252	100.21	Equip.
102	9/16/2015	50	11/4/2015	223	2,185	140.18	Equip.
103	10/4/2015	40	11/12/2015	240	1,719	115.06	Equip.
104	10/13/2015	40	11/21/2015	189	1,803	79.43	Labor
105	11/6/2015	35	12/10/2015	172	1,590	159.43	Equip.
106	11/22/2015	35	12/26/2015	258	1,574	78.80	Equip.
107	12/4/2015	28	12/31/2015	198	1,243	104.27	Manage
108	12/17/2015	30	1/15/2016	235	1,322	99.14	Equip.
109	12/22/2015	50	2/9/2016	230	2,252	113.49	Equip.
110	12/28/2015	35	1/31/2016	179	1,590	88.41	Labor


MACRO IMPACT FACTOR
(PRODUCTIVITY LOSS CAUSED BY)
labor
21%
management
13%
equipment
66%

b. Summary of the Factor Data

Project Data	Impact Factor					Pile Construction Productivity	Total Factor
	Env	Equip	Labor	Material	Mangement		
1	7	10	14	5	11	26	47
2	6	9	10	6	9	21	40
3	5	8	8	7	8	19	36
4	7	7	9	7	10	20	40
5	14	16	20	15	18	41	83
6	16	8	20	19	12	30	75
7	16	6	20	17	12	28	71
8	16	6	20	17	12	28	71
9	16	8	23	20	11	31	78
10	11	10	19	14	7	26	61
11	11	10	18	13	7	26	59
12	12	6	11	16	10	21	55
13	12	8	14	14	8	23	56
14	10	7	12	12	7	20	48
15	12	8	12	12	7	21	51
16	10	6	9	14	9	19	48
17	6	5	8	10	8	16	37
18	9	8	10	9	8	20	44
19	19	20	23	20	19	48	101
20	5	6	11	7	8	18	37
21	10	6	9	10	9	19	44
22	12	9	14	14	9	25	58
23	14	16	20	15	18	41	83
24	7	8	9	7	8	19	39
25	17	10	23	15	12	33	77
26	10	7	12	12	7	20	48
27	16	6	20	17	12	28	71
28	12	9	11	16	10	24	58
29	11	10	16	14	7	25	58
30	12	10	11	11	8	23	52

31	7	7	12	8	8	20	42
32	5	8	8	7	8	19	36
33	7	7	9	7	10	20	40
34	10	8	9	10	9	21	46
35	6	9	10	6	9	21	40
36	11	10	18	13	7	26	59
37	13	8	15	11	9	24	56
38	12	6	11	16	10	21	55
39	6	5	9	10	5	14	35
40	10	8	9	10	9	21	46
41	8	6	12	9	7	19	42
42	8	3	10	9	10	17	40
43	11	9	14	13	9	24	56
44	9	8	10	9	8	20	44
45	14	16	20	15	18	41	83
46	10	6	9	10	9	19	44
47	17	10	23	15	12	33	77
48	11	10	16	14	7	25	58
49	12	10	11	11	8	23	52
50	7	7	9	7	10	20	40
51	10	8	16	10	6	22	50
52	7	5	13	6	4	15	35
53	16	6	20	17	12	28	71
54	11	10	16	14	7	25	58
55	5	8	8	7	8	19	36
56	7	7	9	7	10	20	40
57	11	10	18	13	7	26	59
58	10	7	12	12	7	20	48
59	12	8	14	14	8	23	56
60	6	9	10	6	9	21	40
61	14	16	20	15	18	41	83
62	12	10	11	11	8	23	52
63	16	6	20	17	12	28	71
64	10	7	12	12	7	20	48

65	12	8	12	12	7	21	51
66	16	8	23	20	11	31	78
67	5	8	8	7	8	19	36
68	11	10	16	14	7	25	58
69	12	9	14	14	9	25	58
70	11	10	18	13	7	26	59
71	12	10	11	11	8	23	52
72	5	8	8	7	8	19	36
73	12	10	11	11	8	23	52
74	7	7	12	8	8	20	42
75	5	8	8	7	8	19	36
76	7	7	9	7	10	20	40
77	10	8	9	10	9	21	46
78	6	9	10	6	9	21	40
79	11	10	18	13	7	26	59
80	16	6	20	17	12	28	71
81	9	6	11	11	7	18	44
82	12	10	11	11	8	23	52
83	8	6	12	9	7	19	42
84	8	5	12	9	7	18	41
85	12	8	12	12	7	21	51
86	12	8	15	10	8	23	53
87	13	10	15	14	9	26	61
88	16	12	15	15	12	31	70
89	12	9	11	16	10	24	58
90	12	6	11	16	10	21	55
91	12	10	11	10	8	23	51
92	10	9	12	10	10	24	51
93	14	16	20	15	18	41	83
94	10	6	9	10	9	19	44
95	5	8	8	7	8	19	36
96	12	10	11	11	8	23	52
97	20	8	23	20	15	34	86
98	12	10	11	11	8	23	52

99	12	6	11	16	10	21	55
100	14	16	20	15	18	41	83
101	11	10	16	14	7	25	58
102	11	10	18	13	7	26	59
103	12	10	11	11	8	23	52
104	16	8	23	20	11	31	78
105	14	16	20	15	18	41	83
106	12	8	12	12	7	21	51
107	10	6	9	10	9	19	44
108	9	8	10	9	8	20	44
109	11	10	16	14	7	25	58
110	17	10	23	15	12	33	77

APPENDIX : MODEL RESULT

a. Coding Syntax for Prediction and Productivity Loss

```
library(e1071)
prevCoeff <- 0
finalObj
print("start copy")
datatotal <- read.delim("D:/THESIS/Coding RStudio /data34.txt")
for (datasetx in 35:44) {
  name <- (paste("D:/THESIS/Coding RStudio /data", datasetx, sep = ""))
  namefile <- (paste(name, "txt", sep = "."))
  dataTemporary <- read.delim(namafile)
  datatotal <- rbind(datatotal, dataTemporary)
  print(namafile)
}
for (i in 1:100) {

  model1 <- svm(
    loss~pileDepth+Numberofpile+sizeproject+numberofequipment+location+typeofpilr+soilcondition+city,
    data = datatotal,
    cross = 10
  )
  obj <- summary(model1)
  prediksi1 <- predict(model1, datatotal)
  if (obj["scorrcoeff"] > prevCoeff) {
    #print(obj["scorrcoeff"])
    prevCoeff <- as.numeric(unlist(obj["scorrcoeff"]))
    finalObj <- obj
  }
  # if (obj["scorrcoeff"] > 0.90) {
  #   print(obj)
  #   #break
  # }
}
print(finalObj)
print(prevCoeff)
summary(model1)
prediction1 <- predict(model1, datatotal)
prediction1
print("finish")
```

b. Code Attributes for the factors

Attributes	Code	Code After Normalization	Explanation

Type of soil condition	0	0	Bad Soil Condition
	1	0.14	Peatland
	2	0.29	Clay + Sand (mud/silt soil)
	3	0.43	Rock Soil
	4	0.57	Subgrade Soil
	5	0.71	Clay
	6	0.86	Clay + Sand + Gravel
Type of Pile	0	0.00	Raft foundation
	1	0.17	Prestressed
	2	0.33	Precat piles
	3	0.50	Anchor foundation
	4	0.67	Bored Pile (cast in situ)
	5	0.83	Micropiling or Helical Piling
Location	0	0	Urban
	1	0.5	Sub-Urban
City Location	0	0	Jakarta
	1	0.2	Surabaya
	2	0.4	Malang
	3	0.6	Solo
	4	0.8	Semarang

Project No.	Pile Depth (m)	Number of Pile (quantity)	Size of Project (m2)	Number of Equipment	Location (code)	Soil Condition (code)	Type Of Pile (code)	City (code)
52	13	139	1786	3.03	1	0	0	0
102	23	223	5163	3.14	0	3	3	1
99	21	246	5052	3.10	0	1	1	2
98	21	244	5035	3.10	0	1	1	2
90	21	266	5428	3.10	0	1	1	2
70	23	227	5281	3.06	0	3	3	3
69	21	262	5352	3.10	0	1	1	2
64	22	215	4672	3.06	0	2	2	3
59	21	229	5021	3.03	0	1	1	0
58	22	214	4651	3.06	0	2	2	3
57	24	237	5516	3.14	0	4	4	1
38	21	230	4713	3.10	0	1	1	2
36	24	232	5401	3.14	0	4	4	1
26	32	210	6651	3.06	0	5	5	3
22	21	260	5357	3.10	0	1	1	2
16	24	213	5112	3.29	0	4	4	4
14	22	206	4480	3.06	0	2	2	3
13	22	223	4902	2.95	0	2	2	3
12	21	236	4843	3.10	0	1	1	2

11	13	234	3000	3.14	1	0	0	1
10	33	224	7376	3.14	0	6	0	1

c. Basic Data Sample from 8 Factors Before Normalization

d. Data Input Sample From 8 Factors After Normalization

Project No.	Pile Depth (m)	Number of Pile (quantity)	Size of Project (m2)	Number of Equipment	Location (code)	Soil Condition (code)	Type Of Pile (code)	City (code)
52	0.000	0.515	0.233	0.076	0.50	0.00	0.00	0.00
102	0.503	0.836	0.696	0.181	0.00	0.43	0.50	0.20
99	0.379	0.924	0.681	0.143	0.00	0.14	0.17	0.40
98	0.379	0.916	0.679	0.143	0.00	0.14	0.17	0.40
90	0.379	1.000	0.733	0.143	0.00	0.14	0.17	0.40
70	0.503	0.851	0.712	0.105	0.00	0.43	0.50	0.60
69	0.379	0.985	0.722	0.143	0.00	0.14	0.17	0.40
64	0.437	0.805	0.629	0.105	0.00	0.29	0.33	0.60
59	0.379	0.859	0.677	0.076	0.00	0.14	0.17	0.00
58	0.437	0.802	0.626	0.105	0.00	0.29	0.33	0.60
57	0.527	0.889	0.745	0.181	0.00	0.57	0.67	0.20
38	0.379	0.863	0.634	0.143	0.00	0.14	0.17	0.40
36	0.527	0.870	0.729	0.181	0.00	0.57	0.67	0.20

26	0.926	0.786	0.900	0.105	0.00	0.71	0.83	0.60
22	0.379	0.977	0.723	0.143	0.00	0.14	0.17	0.40
16	0.527	0.798	0.689	0.320	0.00	0.57	0.67	0.80
14	0.437	0.771	0.602	0.105	0.00	0.29	0.33	0.60
13	0.437	0.836	0.660	0.000	0.00	0.29	0.33	0.60
12	0.379	0.885	0.652	0.143	0.00	0.14	0.17	0.40
11	0.000	0.878	0.399	0.181	0.50	0.00	0.00	0.20
10	1.000	0.840	1.000	0.181	0.00	0.86	0.00	0.20

e. Potential Loss Prediction Result from 110 Project

Comparison Data Number	Work Hour Lost in Data Model	Work Hour Lost from Comparison Data	Euclidean Distance
16	76.61	76.65	0.04
53	76.61	76.89	0.28
24	76.61	76.26	0.35
97	76.61	76.18	0.43

60	76.61	76.18	0.43
66	76.61	76.06	0.55
99	76.61	77.76	1.15
4	76.61	77.80	1.20
35	76.61	75.33	1.28
9	76.61	75.24	1.37
72	76.61	78.13	1.52
106	76.61	78.80	2.19
94	76.61	79.05	2.44
77	76.61	79.33	2.72
104	76.61	79.43	2.82
95	76.61	73.09	3.51
85	76.61	80.69	4.08
58	76.61	72.34	4.27
74	76.61	81.73	5.12
27	76.61	71.37	5.24
17	76.61	70.13	6.48
89	76.61	83.20	6.59
90	76.61	85.12	8.51

81	76.61	67.94	8.67
65	76.61	67.87	8.74
15	76.61	87.06	10.45
110	76.61	88.41	11.80
5	76.61	88.42	11.81
3	76.61	64.69	11.91
6	76.61	90.16	13.55
56	76.61	90.93	14.32
33	76.61	91.20	14.59
20	76.61	61.61	15.00
39	76.61	60.99	15.62
80	76.61	60.83	15.78
18	76.61	92.55	15.94
32	76.61	92.78	16.17
55	76.61	92.78	16.17
67	76.61	60.32	16.29
50	76.61	93.87	17.26
40	76.61	94.56	17.95
2	76.61	95.78	19.17

59	76.61	96.13	19.52
34	76.61	96.15	19.54
46	76.61	56.80	19.81
69	76.61	97.15	20.54
8	76.61	55.75	20.86
63	76.61	55.72	20.89
31	76.61	97.61	21.00
62	76.61	97.89	21.28
29	76.61	98.11	21.50
7	76.61	54.87	21.74
44	76.61	99.05	22.44
108	76.61	99.14	22.53
21	76.61	99.28	22.67
42	76.61	99.51	22.90
76	76.61	99.75	23.14
25	76.61	99.77	23.16
101	76.61	100.21	23.60
41	76.61	102.82	26.21
91	76.61	102.92	26.31

75	76.61	49.21	27.40
107	76.61	104.27	27.66
86	76.61	105.82	29.21
30	76.61	106.10	29.49
88	76.61	106.65	30.04
26	76.61	107.47	30.86
78	76.61	45.66	30.95
96	76.61	109.01	32.40
92	76.61	111.29	34.68
48	76.61	112.61	36.00
109	76.61	113.49	36.88
103	76.61	115.06	38.45
38	76.61	116.29	39.68
61	76.61	117.43	40.82
73	76.61	117.64	41.03
54	76.61	119.25	42.65
19	76.61	119.50	42.89
12	76.61	122.22	45.61
68	76.61	123.74	47.14

14	76.61	124.69	48.08
64	76.61	126.36	49.75
71	76.61	128.17	51.56
47	76.61	128.43	51.82
10	76.61	130.22	53.61
49	76.61	130.43	53.82
84	76.61	137.49	60.88
36	76.61	139.01	62.40
83	76.61	139.40	62.79
93	76.61	139.48	62.87
102	76.61	140.18	63.57
82	76.61	140.66	64.05
70	76.61	147.14	70.54
52	76.61	150.88	74.27
23	76.61	153.90	77.29
45	76.61	154.08	77.47
51	76.61	155.88	79.27
105	76.61	159.43	82.82
98	76.61	165.00	88.39

57	76.61	171.91	95.30
100	76.61	172.47	95.86
43	76.61	172.80	96.19
37	76.61	174.76	98.15
28	76.61	175.51	98.90
13	76.61	182.29	105.68
22	76.61	203.25	126.64
87	76.61	209.90	133.29
79	76.61	221.38	144.77
11	76.61	322.40	245.79